universität wien

# DIPLOMARBEIT

Titel der Diplomarbeit

# ASSESSMENT IN CLIL

## A CASE STUDY

Verfasserin

INGRID HÖNIG

angestrebter akademischer Grad

Magistra der Philosophie (Mag.phil.)

Wien, im November 2009

| | |
|---|---|
| Studienkennzahl lt. Studienblatt: | A 343 |
| Studienrichtung lt. Studienblatt: | Anglistik und Amerikanistik |
| Betreuerin: | Ao. Univ.-Prof. Dr. Christiane Dalton-Puffer |

# Table of Contents

**Abstract**

It has been frequently observed in adult education that learners having undergone some kind of CLIL instruction speak more self-confidently and naturally in the 2nd language. In the Austrian context this is usually accounted for by referring to the principle of not assessing CLIL students on the L2. Similar observations have been made elsewhere, but empirical research on assessment is still very scarce. This study was therefore undertaken to explore how assessment is carried out in CLIL, in particular to find out which role L2 language proficiency plays in assessment. My qualitative study was carried out in an upper-secondary grammar school, in History lessons taught in English in two forms. It comprises interviews with four CLIL teachers, observation and video-taping in two CLIL classes, and a teacher's comment on nine transcribed oral exams.

My most significant finding was that, in spite of teachers' assertions to the contrary, they did assess language proficiency and even used it as a major criterion in grading students. As my study further shows, traditional methods of assessment such as oral exams do not do CLIL students' achievements justice. Alternative instruments are required. Devising these, however, would call for formal agreement on language criteria in CLIL education as well as on the weighting of content and language in exams.

In der Erwachsenenbildung wird häufig beobachtet, dass Lerner, die während ihrer Schulzeit CLIL-Unterricht hatten, selbstbewusster und natürlicher sprechen. Sie führen dies darauf zurück, dass an Österreichs Schulen die Fremdsprache in CLIL nicht beurteilt wird. Ähnliche Beobachtungen werden auch anderorts gemacht, aber empirische Untersuchungen über Leistungsbeurteilung in CLIL gibt es noch sehr wenige. Diese Studie wurde gemacht um zu untersuchen, wie Leistung in CLIL wirklich beurteilt wird. Im Besonderen wollte ich herausfinden, welche Rolle die Fremdsprache in der Leistungsbeurteilung spielt. Ich führte eine qualitative Studie in zwei Klassen einer Oberstufe in einem Gymnasium durch, in denen Geschichte auf English unterrichtet wurde. Sie umfasst vier Interviews mit CLIL-Lehrern, Beobachtung von zwei CLIL Klassen mit Videoaufnahmen und dem Kommentar einer Lehrerin über neun transkribierte mündliche Prüfungen.

Das wichtigste Ergebnis aus den Unterrichtsbeobachtungen ist, dass Lehrer trotz gegenteiliger Aussagen die Sprachfertigkeit der Schüler bewerten und diese sogar als Hauptkriterium bei der Notengebung heranziehen. Wie die Studie weiter zeigt, sind traditionelle Methoden der Leistungsbeurteilung, wie zum Beispiel die mündliche Prüfung, für die gerechte Beurteilung der Schüler ungeeignet. Alternative Bewertungsverfahren sind erforderlich, deren Entwurf jedoch eine formale Übereinstimmung von Sprachkriterien im CLIL Unterricht sowie der Gewichtung von Sprache und Inhalt in den Prüfungen voraussetzt.

## 1 Introduction

The rise of Content and Language Integrated Learning (CLIL) is a story of remarkable success and progress. It is a fairly short, but complex story. As a child of the early 1990s CLIL was born into an increasingly integrating and mobile Europe, which led to a need of proficiency in foreign languages. In the intervening years, it has gained ground and has quickly grown in acceptance. At present, CLIL has arrived in almost all European mainstream secondary schools and is frequently included in their curricula to complement formal language teaching.

The prerequisites for the implementation of CLIL are determined by intensifying integration and globalisation processes. The language policies of the EU adapted to the changing circumstances; the objective became to make EU citizens 'plurilingual'. As defined by the European Commission, each citizen should be proficient in three European languages: two languages of the Community and their native tongue[1]. Satisfying this objective poses a challenge indeed to educational systems, which carry the responsibility for providing the teaching of foreign languages to pupils and students. Initiatives to widen the scope of language learning, such as CLIL, were well received. In general, CLIL is not linked with one specific language but is viewed as an educational approach to support linguistic diversity. Yet it comes as no surprise that English is a long way in front in all countries, followed by French and German. The dominant position of English in CLIL is determined and explained by its role as the European lingua franca and its importance as the international language (Dalton-Puffer 2005:41).

The reasons for the implementation of CLIL are cogent and manifold. To begin with, traditional foreign language teaching at school has proved to be unsatisfactory; within their school-life the majority of students does not acquire the competence necessary for effective communication in the foreign languages in their future career. In CLIL lessons students are offered more opportunities for exposure to the foreign language without requiring extra time in the curriculum. Equally significant, CLIL,

[1] European Commission 1996. White Paper. *Teaching and learning: towards the learning society.* Brussels.

which provides content per se, offers students a more natural situation for language acquisition and development than traditional language classes, where the focus is on form and structure, the topics concerning everyday life and culture of the respective language.[2]

In Austria, CLIL started in the late 1980s and early 1990s. Within this period of time efforts to extend language learning fell on fertile ground. In view of the importance of language teaching in Austrian schools, CLIL was welcomed by teachers, parents and students. Austrian schools adopted the concept of CLIL and introduced it into mainstream schools. In the European context upcoming membership in the EU and the opening of Eastern Europe strongly supported its implementation. Simultaneously, domestic political measures promoted increased language education: the 14th amendment to the *Schulorganisationsgesetz* (1993) granted individual schools educational autonomy so as to give them a certain characteristic identity. As a result, many of them have opted for introducing CLIL in order to raise the profile of their concern for language learning.[3]

Against this background, the Austrian Ministry of Education and Culture reacted to the increased interest in this new educational approach and established the "Projektgruppe Englisch als Arbeitssprache" (Project group English as a Medium of Instruction) at the Centre for School Development (Zentrum für Schulentwicklung – ZSE) in Graz in 1992. The group aimed at promoting and integrating CLIL into the curriculum, evaluating new developments in language learning and connecting people working with CLIL (Abuja, 1999:1). Its instigators introduced this innovative approach as "Fremdsprache als Arbeitssprache", (FsAA – Foreign Language as a Working Language) (Nezbeda 2005:7). Due to the dominance of English as the working language, the term "Englisch als Arbeitssprache" (EAA), (English as a Working Language) was adopted and has remained in use among Austrian teachers (oral communication).

There are no general guidelines on when to start with CLIL and how many subjects should be taught in a foreign language in Austrian schools. In which form CLIL

[2] www.clilcompendum.com
[3] http://www.bmukk.gv.at/medienpool/5823/schulrecht4.pdf

teaching starts and how many subjects it incorporates varies from school to school and depends entirely on the respective school's decision.

In recent years there has been a surge of academic interest in CLIL. Research into the theoretical as well as practical implications of CLIL has been carried out in several European countries, in particular in Austria, Germany, Finland and Spain. Assessment in CLIL, however, is still an underexplored area. Among practitioners and educationalists it has repeatedly been referred to as problematic. Uncertainty prevails as to whether language can or may be assessed and, if it is, how it can be accounted for with the traditional modes of assessment.

The development of an approach to assessment with a dual focus on content and language would not only prove beneficial within the CLIL dimension. It might also serve well in larger frameworks of education as it is the learning reality for a growing number of children to be instructed in a language other than their mother tongue (e.g. "immersion", immigrant education, science through EFL). (Mohan at a lecture at Vienna University, December 2007). In view of this, issues related to the role of language in assessment are becoming increasingly significant.

## 1.1 Defining the problem

Despite the burgeoning growth of research in CLIL from different perspectives, the issue of assessment in CLIL has hardly been investigated. Two areas, in particular, seem to invite investigation: the role that the language plays in assessment and the appropriateness of traditional assessment tools:

> Im bilingualen Unterricht spielt […] die Leistungsfeststellung bzw. die Leistungsbewertung eine spezifische und zentrale Rolle. Dabei interessiert vor allem die Frage, ob bzw. in welcher Form die sprachliche Leistung einen Einfluss auf die Bewertung der Gesamtleistung im Sachfach haben darf. Die Entwicklung spezifischer Bewertungsverfahren stellt in diesem Zusammenhang ein Desideratum dar (Helbig 2003:184).
>
> *In bilingual education […] assessment plays a specific and central role. Thereby it is of particular interest if and in which form linguistic achievement may influence the overall assessment of the content subject. In this context the development of specific methods of assessment are a desideratum.* (*My* translation)

The growing demand for appropriate means of assessment has not passed unnoticed but has been addressed and reflected on by practitioners as well as educationalists. They have repeatedly voiced their concern about the inadequacy of traditional means of assessment measuring students' achievements in CLIL. At workshops in Graz, for instance, teachers raised the problems and pointed out their needs for specific training and their particular interest in the role of language in assessment (Grangl 1998:192).

> Heiße Diskussionen gab es immer wieder zum Thema Leistungsbeurteilung:
>
> - Soll das Themengebiet auch oder fast ausschließlich in englischer Sprache wiederholt oder abgefragt werden?
> - Wie schlägt sich der Rückgriff auf die Muttersprache auf Punktewertung oder Benotung nieder?
>
> *// There were ongoing lively discussions about assessment:*
>
> - *Should content be revised and examined also or only in the English language?*
> - *How does code-switching influence assessment?//*

When CLIL was still in its early days, Ernst (1995: 258) criticised the lack of publications about correction and, in particular, about assessment in CLIL. Schmid-Schönbein and Sigismund (1998:204) state that the problems with assessment in CLIL have not yet been solved; Vollmer takes the same line when arguing that assessment in CLIL is still a relatively unexplored field.

> Dennoch wissen wir über die Praxis der Leistungsfeststellung und –bewertung im bilingualen Unterricht[4] so gut wie gar nichts […] (2001:207).
>
> *// However, we know next to nothing about assessment in bilingual lessons //*

Koch (2002:84-96) discusses the complex problems of assessment in CLIL in detail and identifies the development of effective assessment procedures as a matter of urgency. In particular, he recommends that innovative approaches to assessment

[4] In Germany, CLIL is referred to as ‚bilingualer Sachfachunterricht'.

should be dealt with in teacher training at universities as well as discussed among colleagues in schools.

Fehling (Finkeiner and Fehling 2002:22-32) conducted a small-scale survey on how CLIL teachers assess written exams. Its findings echo in many ways the problematic issues of and major concerns about assessment (see chapter 3.4).

## 1.2 Defining the research questions

In this study, I aim at identifying the problems addressed and looking at suggested solutions in the literature available. In the empirical part of this study, I investigate how assessment in CLIL is carried out in practice, placing particular emphasis on the role played by the language in assessment.

The specific research questions are:

1. What are the teachers' main concerns that are voiced in the literature
2. Which types of assessment are applied in CLIL
3. Which role does the target language play in assessment, in particular, does linguistic achievement influence the grade
4. Are traditional modes of assessment adequate to cater for the dual focus in CLIL

## 1.3 The structure of the study

The following chapter presents a selective review of the legal regulations and the components of assessment considered against the background of CLIL instruction in the Austrian classroom. As the oral exam turned out to be the most common mode of assessment in my investigation, it is discussed in detail. Additionally, the chapter touches upon the need for change in assessment and exemplifies the request for alternative methods with a brief description of portfolio work.

Chapter 3 shifts the focus from general features of assessment towards CLIL-specific issues. It gives an overview of the tension between content and language in

assessment and summarises practice of and suggestions for error correction in CLIL. Furthermore, it presents portfolio work as a model of alternative assessment in CLIL.

Chapter 4 discusses the role of language as a medium of learning. It looks at the construction of knowledge in and methodologies suggested for the CLIL classroom since assessment cannot be viewed in isolation but is strongly influenced by what and how teachers teach and students learn.

Chapter 5 introduces the empirical part of the study by describing the access to the field, the collection of data and the institutions and teachers, who participated in the data gathering.

Chapter 6 examines assessment in the Austrian CLIL classroom. It outlines the teachers' attitudes towards CLIL, investigates the components of assessment as well as the choice of code and error correction. Again, it is the oral exam, being most frequently applied in CLIL, which is analysed in depth. Additionally, and perhaps more importantly, the chapter highlights the controversies about the role the language plays in assessment between the teachers' perception and actual conduct in the classroom.

Chapter 7 presents the summary of and reflection on the findings and suggests possible solutions for giving oral exams more validity. It briefly describes how well CLIL worked in the classroom observed despite the large discrepancies noted. It also delineates the procedure I had to follow in order to get tangible results.

Chapter 8 concludes the study with an invitation to the parties involved in this innovative approach for the development of assessment strategies that cater for both language and content in assessment.

# 2. Assessment

## 2.1 The nature and purpose of assessment

It is everyday reality for classroom teachers to assess their students' achievements in the subjects they teach. They do it constantly on an individual basis to gain information about their students' progress as well as reserve time during the lessons,

usually after a learning unit, to set oral or written exams. Both strategies allow the teachers to judge their students' knowledge and skills systematically for the purpose of gathering evidence for grading.

Assessment is referred to as "the process whereby one person (usually a teacher or examiner) attempts to find out about the knowledge, attitudes or skills possessed by another (learner). This may involve him in merely observing the students as he sets about normal learning activities; or he may need to create special assessment activities, e.g. quizzes, exams, oral tests, etc." (Rowntree 1981:14).

Based on this description assessment can be classified in various ways[5]. A distinction is often drawn between:

- formal and informal assessment
- formative and summative assessment
- holistic and analytic assessment

*Formal - informal assessment*

Formal assessment normally means the use of tests and exams to assess learners' knowledge or ability. Informal assessment usually refers to a collection of information by means of observation, student's participation or discussion in class (Rowntree 1981:94).

*Formative - summative assessment*

Formative assessment is usually carried out throughout a term or course or project. Ideally, it supports the process of learning and is advice for or feedback on students' work rather than used for grading purposes. It can be gathered from formal and informal assessment in the classroom and serves to inform teachers on how to improve their teaching as well as students on how to improve their learning (Rowntree 1981:95, Jürgens 2000:73).

[5] These are the classifications used in CLIL contexts; for a more extensive list see Rowntree 1981.

Summative assessment is usually carried out at the end of a term or course or project and is typically used to assign students a grade. It sums up formal or informal assessment with the purpose of reporting achievement to students, teachers, parents, school boards (Rowntree 1981:305, Jürgens 2000:73).

*Holistic - analytic assessment*

Holistic assessment addresses the student's whole performance, not just aspects of it. Analytic assessment addresses the student's performance on independent aspects of the task, not on the task as a whole (McNamara 1996:43).

There are various functions that assessment is required to fulfil. These are basically divided into social and pedagogical functions. The social functions serve as a selection criterion, permitting or denying children access to continuing education. The pedagogical functions provide feedback to students to improve their learning, to teachers to reflect on their teaching, and to parents to get informed about their children's achievements and limitations. What the latter have in common is that they serve to raise achievement and improve learning (Jäger 2000:32-39).

In recent years, the notion and purpose of assessment have been redefined, emphasising its pedagogical functions: "the focus [in formative assessment] is on the role of the student and on interaction" and it is "viewed as an integral part of teaching and the development of learning opportunities" (Kiely 2009:2). Because of its support for learning it is referred to as "assessment for learning". Summative assessment, on the other hand, serves to make judgements about standards of achievement and is defined as "assessment of learning" (Kiely 2009:2).

In this context it is perhaps necessary to note that grades not only reflect the students' achievements but also express the quality of teaching (Neuweg 2006:8,122). That is to say that the knowledge and skills children can acquire at school are closely linked to the individual teacher's capacity, effort and commitment. Furthermore, the teacher's behaviour during the exam can considerably influence its results, particularly in oral exams (Neuweg 2006:8, cf. chapter 2.5).

## 2.2 Legal regulations

The legal guidelines for students' assessment in Austrian schools are provided in The School Act[6], in particular in §18 - 23, and in the act for assessment (Leistungsbeurteilungs-verordnung) and its amendments[7]. The legislator distinguishes between two phases in assessment (Neuweg 2006:7-8):

1. Leistungsfeststellung; the teacher's collection of evidence of learning results, based on, for instance, the number of correctly answered questions, the number of errors in the tasks or the number of points the student has achieved in an exam.

2. Leistungsbeurteilung: the teacher's judgement of the evidence against a norm and its final expression in a numerical grade.

Although assessment can be practiced without any kind of measurement, there is usually no distinction drawn between the collection of the learning results and the subsequent grading in our educational culture. In practice, these two steps normally merge into one. Winter even speaks of a mathematical equation: assessment = grading (2004:312).

### Grades

Grades are defined as the "shortform of the teacher's judgement about the student's achievement" (cf. Schaub und Zenke 2007:706). Since their introduction in the early 19th century grades have been a bone of contention. While historically assessment in the form of numerical grades has been used for selection, its use has continued because it is considered to be based on principles such as objectivity, reliability and validity (see chapter 2.5) (Winter 2004:41). Particular criticism has been levelled against the objectivity and validity of grades: different teachers tend to grade achievements differently, depending on the norms and criteria they use and their judgements are often affected by errors (see below).

[6] http://www.bmukk.gv.at/schulen/recht/gvo/schug_teil1.xml#18
[7] http://www.bmukk.gv.at/schulen/recht/gvo/lb_vo.xml#01

## The Austrian grading system[8]

In primary, secondary and tertiary education the grading system is based upon a five-point scale with 1 as the highest achievable grade and 5 as the lowest:

Each grade is provided with a definition that lays down the criteria for assessment. The definitions for the grades 1 to 5 are spelled out as follows:

> Mit „Sehr gut" sind Leistungen zu beurteilen, mit denen der Schüler die nach Maßgabe des Lehrplanes gestellten Anforderungen in der Erfassung und in der Anwendung des Lehrstoffes sowie in der Durchführung der Aufgaben in weit über das Wesentliche hinausgehendem Ausmaß erfüllt und, wo dies möglich ist, deutliche Eigenständigkeit beziehungsweise die Fähigkeit zur selbständigen Anwendung seines Wissens und Könnens auf für ihn neuartige Aufgaben zeigt.
>
> Mit „Gut" sind Leistungen zu beurteilen, mit denen der Schüler die nach Maßgabe des Lehrplanes gestellten Anforderungen in der Erfassung und in der Anwendung des Lehrstoffes sowie in der Durchführung der Aufgaben in über das Wesentliche hinausgehendem Ausmaß erfüllt und, wo dies möglich ist, merkliche Ansätze zur Eigenständigkeit beziehungsweise bei entsprechender Anleitung die Fähigkeit zur Anwendung seines Wissens und Könnens auf für ihn neuartige Aufgaben zeigt.
>
> Mit „Befriedigend" sind Leistungen zu beurteilen, mit denen der Schüler die nach Maßgabe des Lehrplanes gestellten Anforderungen in der Erfassung und in der Anwendung des Lehrstoffes sowie in der Durchführung der Aufgaben in den wesentlichen Bereichen zur Gänze erfüllt; dabei werden Mängel in der Durchführung durch merkliche Ansätze zur Eigenständigkeit ausgeglichen.
>
> Mit „Genügend" sind Leistungen zu beurteilen, mit denen der Schüler die nach Maßgabe des Lehrplanes gestellten Anforderungen in der Erfassung und in der Anwendung des Lehrstoffes sowie in der Durchführung der Aufgaben in den wesentlichen Bereichen überwiegend erfüllt.
>
> Mit „Nicht genügend" sind Leistungen zu beurteilen, mit denen der Schüler nicht einmal alle Erfordernisse für die Beurteilung mit „Genügend" (Abs. 5) erfüllt.

[8] Verordnung über die Leistungsbeurteilung in Pflichtschulen sowie mittleren und höheren Schulen (Leistungsbeurteilungsverordnung) §14.

The legislation distinguishes two aspects that are to be expressed in the grades: the reproductive aspect, namely comprehension and use of what the students have learned and, secondly, the productive aspect in form of the independent use of their knowledge and their ability to use it for new tasks (Neuweg 2006:80).

Assessment is not an abstract process but inevitably incorporates particular values; these are specified in comparison to a norm. In the educational process students' achievements are assessed against three reference norms:

- curricular norm
- social norm
- individual norm

*Curricular norm*

For the purpose of equality and fairness the school law allows exclusively the curricular norm as a valid standard for measuring students' achievements (§11 LBVO - "Maßstab für die Leistungsbeurteilung sind die Forderungen des Lehrplanes unter Bedachtnahme auf den jeweiligen Stand des Unterrichts"). The student is compared against a pre-determined standard, which is defined by the learning objectives of the curriculum. What is assessed is to what extent the student has reached these learning objectives. (Neuweg 2006:86-87).

*Social norm*

The student is measured in rank order in comparison to his peers in the same classroom. The teacher works out the average of the achievements and then grades the individual student in relation to this average. Typically, this norm is not defined before the teacher corrects the exams. Assigning grades according to the social norm is commonly practised at school and tacitly tolerated although it is in direct contrast to the school law. The grade the students get does not reveal how close they have come to matching the learning objectives (Neuweg 2006:83-86, 75-76).

*Individual norm*

The individual norm serves to foster the students' development. Their achievement is measured against their individual potentials, for instance, by comparing the students' progress between present performance and prior performance. While theoretically

this might prove beneficial for the students, in practice it negates the functions of assessment, as it does not tell the distance of the student's knowledge to the expected curricular standard. Therefore it has been suggested to combine the individual norm with the curricular norm by assessing the students against the curricular criteria but emphasising their individual progress (Neuweg 2006:88-89).

In the assessment literature the reference norms are commonly referred to as problematic. They attach grades a certain value only within the classroom but do not allow grades as a basis for comparison across student cohorts, even between two forms at the same level and subject at the same school. Judging students against the curricular norm might suggest the possibility of comparison as the curriculum establishes the criteria for each individual subject (see above). There are, however, no clear and precise definitions that explicitly tell what students need to know in a particular subject at a certain level at school to achieve a certain grade (Rheinberg 2002:66-67).

In view of the problematic issue of norm-referenced assessment it might be interesting to look at how assessment in Austrian schools is seen from "outside". Educational experts from England, being accustomed to external examinations, feel puzzled about the "degree of freedom that teachers/lecturers enjoy in the Austrian system of testing and evaluation" (Häusler-Greenfield,1999:4:83). She then raises the question "What exactly are the arguments against having uniform agreed procedures, generally accepted criteria or even can-do statements (descriptors)?" Morgan (2006:61) addresses this matter in a somewhat euphemistic way by talking about Austrian teachers having "far more flexibility and more possibility of exploring radically different assessment procedures". As someone educated and practising within the Austrian education system, I am used to the "freedom we enjoy". I do however deplore the reluctance of some Austrian practitioners to explore "radically different assessment procedures".

### 2. 3 Learning goals and objectives

Marks and grades are only the final product of assessment. Therefore, one further and crucial aspect of assessment that needs to be attended to is the answer to the question when assessment begins. The necessary, and indispensably, first step in

designing an assessment concept is to define learning goals and objectives. In other words, the process needs to begin with the goals and objectives, not the methods, modes and tools. Only when these are clearly articulated, can appropriate modes and tools of assessment be selected.

There is an essential distinction between learning goals and objectives, the difference being the level of specificity. Both will vary according to specific contexts but nonetheless serve as a useful guide as to which competencies and abilities should be targeted (Mager 1977:3).

*Learning goals*

Goals are generally broad and in most cases difficult to measure but serve as a means to focus on a big and important issue. They are the broad outcomes that are expected. These general educational goals are to be translated and broken down into specific learning objectives.

*Learning objectives*

Learning objectives, on the other hand, are statements in specific and measurable terms that describe what a learner will know or be able to do as a result of engaging in a learning activity. To find out if students have learned what was intended, results are measured, and it can be determined if the objectives have been achieved. When there are no clearly and specifically defined objectives, there is no sound basis for the selection of instructional materials, content or method; neither is there a sound basis for effective modes and tools of assessment. "Wenn Sie nicht wissen, wohin Sie wollen, ist es schwer, geeignete Mittel auszuwählen, um dorthin zu gelangen"(Mager 1977:5). (*If you do not know where you are going, it is difficult to select suitable means for getting there.* My translation.) Objectives describe the desired outcomes of learning and therefore form a solid basis for assessment.

*The design of learning objectives*

The necessary prerequisite for the defining of learning objectives is the needs analysis (or needs assessment). Two questions have to be dealt with and decided. In the first place it is necessary to check if there is a reason to learn something, and second, to make sure if the learners do not already know what one intends to teach (Mager 1977:1). Needs analysis has to be carried out before objectives can be

designed. Only then can the next step to devising the learning objectives be taken. Once these learning objectives are in place, the routes towards them can be modelled and learning activities planned.

Before embarking on the process of designing the learning objectives teachers should be clear about what they want their students to achieve. The curriculum of the subject provides the framework for the learning objectives; which questions the teacher assumes as relevant to ask is an individual decision. It is suggested that teachers set up a catalogue of questions for particular subjects and levels in co-operation with colleagues of the same subject and establish criteria of what students should know and be able to do. It can be assumed that teachers are concerned that their students understand the "essentials" and are able to use their knowledge creatively in different contexts. In practice, however, teachers gravitate towards factual questions as these tend to be simpler to state, easier to answer and less controversial when correct answers are announced (Neuweg 2006: 15-16).

## 2.4 Learning goals in CLIL

With respect to the language dimension in CLIL, the first question regarding needs analysis, namely if there is a reason to learn something, actually, seems to be unnecessary as it would call into question CLIL instruction as well as language support in the CLIL classroom. As to the second question if the learners do not already know what one intends to teach: CLIL was implemented to enhance learners' language competence and it can be assumed that learners do need language support (Thürmann 2000:87). As a matter of fact, however, there is no explicit curriculum for CLIL classes, featuring concrete language learning goals (Dalton-Puffer 2005:10).

The *CLIL Compendium*[9] presents the following three language-specific learning goals:

- Improve overall target language competence
- Develop oral communication skills

[9] www.clilcompendioum.com

- Deepen awareness of both mother tongue and target language

The language-specific potentials as expressed by Austrian stakeholders are rather unspecific; the goals formulated are "increasing exposure, increasing practice, increasing language competence" (Dalton-Puffer 2008:2). If we look at these goals, it comes as no surprise that they cannot be achieved in their entirety in the curriculum. The language goals in CLIL, addressed in the *CLIL Compendium* as well as being expressed by stakeholders, are too generally formulated to be regarded as objectives. The crux of the matter, it can be assumed, is the lack of concrete language objectives and might perhaps answer the question as to why assessment has remained a problem in CLIL. The absence of language-specific learning objectives makes it difficult, if not impossible, to have the assessment of language under control. The foreign language is not the teaching subject but the working language and the medium of communication. Therefore the learning goals of the content subjects are in the foreground. For the teacher, it is common practice to define content-learning objectives in the curriculum. However, to make CLIL teaching effective with respect to language improvement, language-learning objectives need to be integrated in the curriculum. In other words, CLIL necessitates carefully specified content- as well as language-learning objectives as it is unlikely that the desired language proficiency will be reached without due attention being paid to this. Otherwise, the opportunities provided by CLIL will be missed and the 'value added' significantly reduced.

## 2.5 Components of assessment

The following list comprises the components of assessment that I found applied in CLIL in my investigation. They are included in the law for assessment (Leistungsbeurteilungsverordnung) §3.[10]

- students' participation in class
- oral exams
- written exams

[10] http://www.bmukk.gv.at/schulen/recht/gvo/lb_vo.xml#01

The legislator attaches the same value to all the components of assessment. Yet the practice in the classroom does not always comply with these legal regulations. So it is that particularly written but also oral exams are usually given more significance than participation in class (Neuweg 2006:29).

The advantages and disadvantages of every component of assessment are usually discussed in comparison to each other in their relation to objectivity, reliability and validity. These three criteria provide the framework which makes it possible to judge the quality of measurements and to estimate the effects of measurement errors.

**Objectivity** refers to the requirement that different teachers should come to the same judgement of a student's achievement (Ingenkamp 1997:34, Ziegenspeck 1999:133). Additionally, objectivity is the precondition for the realisation of reliability and validity (Jäger 2000:186, 187). The representation of achievement in form of numerical grades suggests objectivity (Olechowsky und Persy 1987:31). However, different teachers tend to assess the same achievement differently.

**Reliability** indicates the consistency of judgement across teachers and over time, which measures the same thing (Ingenkamp 1997:38). If a particular test is administered several times to the same student, the results should be the same. In practice, reliability is inconsistent and decreases considerably over a longer period (Ingenkamp 1997:38).

**Validity** relates to the extent to which a test measures what it is intended to measure. It holds when the assessment procedure measures the performance described in the objective, that is what it claims to measure (Ingenkamp 1997:42). However, there are often wide discrepancies between the learning objectives and the questions asked in an exam: for example, if students' understanding of literature should be enhanced, they are set an essay to write on literature (Olechowsky 1987:31,32). Moreover, teachers should carefully consider which factors might influence the perception of a the student's performance, for example their linguistic skills when assessing their content knowledge or their ability to memorise when asked to demonstrate the understanding of the content.

**Participation in class**

Participation in class encompasses all the student's achievements during the lessons such as revisions, active contributions in developing new topics as well as home-exercises if required. Evidence of students' participation needs to be used in each subject in order to balance oral and written exams. In those subjects where there are no other exams, oral or written, the appraisal of students' participation forms the basis for summative assessment. All the work performed by the student, whether autonomously or in partner- or group work, should be taken into account, but should be assessed individually (Neuweg 2006:29-32).

**Oral exams**

In my investigation the majority of exams taken in CLIL classes are in oral form. Therefore I will discuss these in more detail.

An oral exam can be defined as the performance of a student's achievement towards a teacher or assessor in form of spoken questions and answers. Needless to say that the performance in an oral exam is inextricably bound up with language. The oral exam is used for the evaluation and prognosis of achievements in nearly all fields of education and covers a broad spectrum of assessment (Jäger 2000:174-175).

In comparison to written exams, oral exams have certain advantages such as immediate feedback and individual adjustment to the student's level of competence. However, they are time-consuming because they require each student to be assessed individually. Furthermore, and more importantly, it is agreed that oral exams severely lack in objectivity, reliability and validity (Neuweg 2006:44-45).

- In oral exams the questions differ whereas in written exams all students get the same questions. Moreover, the conditions under which the students are examined are not the same. If the teacher is perceived by the student as encouraging and fair, the student's attainment is better. Additionally, the sequence of questions influences the student's performance. Ideally, the teacher should begin with easy, so-called "icebreaker" questions and then proceed with more difficult questions.

- Oral exams are “fleeting”. The teacher must simultaneously listen to the student and judge their performance. She must decide very quickly without having the possibility to check the answers a second time.
- Oral exams are less valid because the proportion of speech (Sprechanteil) and its speed (see below) decide on the grade. Moreover, students that are perceived by the teacher as calm, relaxed and concentrated or are not too difficult get better grades.
- Some students experience a high level of stress in oral exams.

Doing well in oral exams requires different skills than those asked for in written ones. In both, students need mastery of the content but the access to and articulation of the knowledge is different. That is why some students enjoy oral exams and others hate them; they are favoured by eloquent students but perceived as stressful by those not so articulate ((Jäger 2000:178).

The factors constituting the achievement in performance are, next to knowledge of content, the following (Jäger 2000:175):

linguistic skills:
pronunciation
expression
flow of argumentation
style
rhetoric
fluency

From the point of view of communication psychology the “negative” factors influencing oral exams can be explained if an oral exam is seen as a relation between sender (student) and receiver (teacher, assessor). Both student and teacher send information that is influenced by additional factors, which explains why the information might arrive at the receiver in a modified or reduced version. These factors are:

- The communicative factors such as facial expression, gesture or rhetoric.
- The personal factor, which tells something about both student and teacher.

- The relationship factor, which communicates the atmosphere between student and teacher, if, for instance, one likes the other or not or if teachers prefer some students to others (Jäger 2000:201-203)

That is to say that emotional processes always coincide with assessment of oral exams, but in most cases they are neither perceived by the teacher nor by the student (Jürgens 2000:84).

Additional critical factors that are likely to modify the results of a student's oral performance are the following (Jäger 2000:182/201):

- The same student is judged differently by different teachers.
- Teachers with long experience in both teaching and assessing tend to be stricter in judgements
- The impression the teacher gains of the student influences the judgement.
- Higher speed in performance is most likely associated with higher competence and abilities.

Pritz (1981, quoted in Jäger 2000:198/199, Neuweg 2006:45) demonstrated that the speed of student's performance influences the grade the student could achieve. He videotaped an oral school leaving exam in Geography, which was graded 3. Then the same student repeated her performance twice, which was again videotaped. Once in a slow version of 23 minutes, the student spoke hesitatingly with little pauses, yet without stuttering; the second time speaking faster and more fluently with firm conviction in 16 minutes. 81 teachers randomly judged either the slower or faster version. The teachers' average grade for the slower version was 3,38 while the faster was graded 2,51 on average. Pritz's study gives substance to the assumption that the more fluently and faster a student speaks and the more proficient they are in the language, the better they can communicate their knowledge. And it is on this basis where teachers partly award their grades to (Jäger 2000:199).

A further influencing factor has been identified with regard to the sequence in which several students' performances are considered. This is called the serial effect. Birkel (1987, quoted in Jäger 2000:197) showed that students are compared with each other

in exams, more precisely, the performance judgement of a student is influenced by the performance of the prior student. Birkel videotaped an oral school leaving exam of two students in the subject German, which was then shown to 156 teachers twice, the sequence of the students' performances being changed.

In comparison with each other the grades the students achieved were skewed in the following way:

- The better student was assigned a better grade when her performance followed the weaker student.
- The weaker student received a lower grade when she followed the better student.

The serial effect has also been shown to exist in written exams as the performances of different students are also judged in a sequence (Jäger 2000:198).

Despite the many limitations discussed, oral exams also offer a number of advantages (Neuweg 2006:46):

- In real life situations oral proficiency and contact play an important role.
- Skills such as improvisation, spontaneity or ability to adapt in a dialogue can only be instantiated in an oral exam.
- Oral exams are more flexible in terms of the level of difficulty. Teachers can adapt to the student's level of competence or can prompt them if they get stuck, whereas students can ask if they have not understood the question.
- Preparing and assessing oral exams takes little time (provided the number of examinees in class is low).

Given the usefulness of oral exams despite their limitations, suggestions have been made to improve the validity of oral exams and assess what is intended to be assessed (Neuweg 2006:47/48).

- First and foremost, it is recommended that teachers prepare the questions and write them down prior to the exam.

- The number of questions should be defined prior to the exam and should be adhered to so that students with a different speed can get equal chances.
- The student should be given appropriate time to think about the question.
- The teacher should be aware of the influence of the serial effect as well as the speed of delivery.
- Each answer should be assessed individually and noted down; otherwise the global impression might decide the grade.
- The grade should be told the student immediately after the exam.

**Written exams**

The most common forms of written exams are essays or a set of questions that require short or more complex answers (Neuweg 2006:51/58). In CLIL classes normally the latter is put to use in exams.

Compared to oral exams, written exams are attributed more objectivity since all students take the exam under the same conditions; they are presented with the same questions of the same number in the same sequence within the same time limit (Neuweg 2006:44). Most students feel more at ease with written exams as they can organise their own time and their sequence of answering; they can go back to questions they missed out or think they might have got wrong.

Teachers assess their students' knowledge using pre-set criteria, that is to say that each question has a known correct answer. It is suggested that a complex question is split into single elements, so that each correctly answered element can gain points. The number of points the students can achieve for the total exam and for each single question and its elements should be laid down prior to the exam. For the questions or single elements that remain unanswered the teacher then should deduct points (Neuweg 2006:71).

In written work it is easier to make selective correction. In this case, the teacher decides to correct only certain errors; which errors will be corrected is usually decided by the objectives of the individual subject.

But also the results of written exams are of doubtful validity as the following criteria flow into the judgement and might skew the grade (Jäger 2000:229-230):

- longer essays achieve better grades
- bad handwriting influences the result negatively
- grammatical and spelling errors have negative effects on the assessment

Furthermore, some factors that might alter the grade in oral exams (see above) can also affect written exams (Jäger 2000:231):

- the serial effect
- the relationship factor
- teachers' experience

For the students' benefit the combination of written and oral exams is suggested as some students are better at speaking, others at writing. Furthermore, students will be likely to encounter both modes in their further academic and professional lives.

## 2.6 A call for change

Important decisions are made on the basis of assessment; and yet for all this, assessment had long been the most neglected element in curriculum planning. In recent years it has gained in attention. Changing views on teaching and learning in combination with new demands in society on students' knowledge and competences call for major educational reforms (Winter 2004:1). The need for a change is underlined by Austrian's pupils' performance in comparative studies such as PISA (Programme for International Student Assessment) or TIMMS (Third International Mathematical and Science Study).

Winter makes the point that intended reforms in education can only be effective if assessment is being changed simultaneously. Since assessment is one of the most important influences on how teachers teach and students learn, improvement efforts that exclude assessment are bound to fail (2004:1). Whereas assessment has traditionally been used primarily for selection and certification, Winter calls for widening its purposes, foregrounding the pedagogical functions (see chapter 2.1).

Assessment should not only be carried out at the end of a learning unit for the purpose of gathering grades but also needs to be integrated into the learning process as a tool to support teaching and learning (Winter 2004:15). Ideally, it should provide information for the student and not about him (Klafki 1996:234, quoted in Winter 2004:95). In other words, the role of assessment needs to be redefined from assessment of learning to assessment for learning (cf. chapter 2.1).

New assessment procedures and tools are required to suit the changing needs. In recent years, educational reformers have recommended portfolio work. Indeed, it seems to be the panacea that satisfies these needs.

**Portfolios**

Portfolios are both old and new. They have long been in use in the visual and performing arts, in which personally favoured works and accomplishments are collected and presented. The selected works, and not other people's judgement about them, serve as evidence of competence and abilities. In the educational process they have been put to use only recently and are slowly gaining a foothold. They can be implemented with the purpose of reforming teaching as well as assessing.

In the context of education a portfolio is beyond the mere collection of what students consider their best-accomplished tasks but "is a purposeful collection of student work that exhibits the student's efforts, progress or achievement in one or more areas. The collection must include student participation in selecting contents, the criteria for judging merit, and evidence of the student' self- reflection" (Paulson, Paulson&Meyer 1991:60, quoted in Jäger 2000:288).

Portfolios are applied in language learning in the form of the "European Language Portfolio"[11] based on the Common European Framework of Reference, with its detailed criteria for self-assessment and competence in a foreign language. In all curricular areas they are gradually being introduced. Here they are flexible tools for both instruction and assessment, which may centre on one subject or span two or more curricular areas. They come in different sizes and shapes, serve various purposes and can be adapted to diverse curricula and all levels. What they all have in common are the stages and processes the students have to go through: collection,

[11] http://www.coe.int/T/DG4/Portfolio/?M=/main_pages/levels.html

selection, connection, reflection and the presentation to an audience. At the core of portfolio work is the student's self-reflection. The students are required to explain and comment why they have chosen specific information, why they link it in a particular way, and how they manage their work, which makes them aware of the learning processes and the methods they have to apply to accomplish their tasks. Thus they develop transferable competences, which enable them to become life-long learners.

Portfolio work needs thorough planning, aligned with the curriculum and fostering students' learning process. Before getting started on their work, the students need clear instructions from their teachers as to what they are expected to do for their portfolio work, and its final presentation. Teachers set objectives in order to then determine and assess the students' progress and achievement in relation to these objectives. They then depart from their traditional role as instructors and adopt the role of counsellors or coaches, who guide and support their student through the processes of their work. The students in turn take over control and responsibility for their learning. The level of students is significant in the process of guidance: in lower classes the students will need repeated help of the teacher while students of the upper school may accomplish their portfolio on their own. Older students may also negotiate the objectives with their teachers and may assess their portfolio work.

Pfeifer highlights the benefits of working with portfolios in a foreign language. She points out that students not only acquire content-specific vocabulary but, more importantly, they enhance their communication skills in the foreign language when they exchange information with their peers or give presentations on their topic to an audience (2007:100; cf. chapter 3.4).

Portfolio work is a promising strategy for the enhancement of instruction and assessment. Nevertheless, it comes at a price. Portfolios require additional skills and extra time from teachers for planning, developing strategies, interacting with students and revising and commenting on students' work. But the burden seems worthwhile. In addition to the above mentioned benefits, portfolio work increases motivation, which is a powerful impetus for learning. A pleasant side effect of portfolio work is that it creates an atmosphere of active co-operation in class. Students learn how to support each other and how to give encouraging and constructive feedback to their

peers. (Pfeifer 2007:41). Both, lack of motivation and the competitive nature of the educational system our students are accustomed to are central issues in the ongoing debate on educational reforms in our country.

Adherents claim that portfolio work proves beneficial for both teachers and students. For teachers it gives evidence of students' achievements and their learning progress over time in given areas. Its presentation offers them the opportunity to communicate about their students' achievements with colleagues, which might instigate co-operation for the sake of quality in education (Winter 2004:139-140, 313). The positive effects of portfolios on students are that they become actively involved in learning and assessment; they become aware of the whole process of learning and which skills they have to use to reach their goals. Self-assessment in portfolio work fosters their self-control, self-criticism and self-responsibility (Jäger 2000:305-309).

Portfolio work has been promoted since the 1970s (Vierlinger 1999), but it has been met with little response, remaining a minority pursuit in education. There is one strong (?) obstacle to its implementation into the learning process. It is a difficult challenge for teachers as they are firmly rooted in traditional teaching and assessment procedures and not too willing to change them (Winter 2004:313; cf. chapter 7.1).

# 3 Assessment in CLIL

## 3.1 The tension between content and language in assessment

Assessment in CLIL is a complex matter. Within the CLIL approach, it is probably the most contentious element and poses a host of challenging questions. Should teachers assess content knowledge and language competence; should intercultural competence be taken into account, and if so, how can achievement be identified and measured quantitatively? However, teachers' concern about assessment in CLIL does not so much focus on the content of the subject taught as its objectives are defined. The bone of contention is the role of language; the key question repeatedly raised is if language skills should be part of assessment, and if they are, how they can be assessed.

McNamara argues that of all the questions asked about assessment, the most important ones are: who wants to be informed about what? (personal communication). Yet answering these questions satisfactorily in matters of CLIL is not as straightforward as it might seem. Only the first part of the question, referring to the parties to be informed, can be answered clearly and quickly: the students, the teachers, the school, and the parents.

The second part, namely what the above-mentioned parties want to be informed about, is relatively opaque and needs special consideration. The literature alludes to the difficulties, indicating the problem of deciding whether to appraise or assess content and language independently.

> Die Konsequenzen für eine erweiterte wie integrative Leistungsmessung sind allerdings noch nicht annähernd durchdacht, vor allem im Hinblick auf die Frage, ob fremdsprachliche und inhaltliche Leistungen, so verschränkt sie ja sind, getrennt überprüft oder gar bewertet werden sollen (Vollmer 2001:215). *The consequences of an extended such as integrated assessment have hardly been considered, particularly in view of the question if achievements in the foreign language and content, although intricately interwoven, should be appraised or even assessed separately* (My translation).

The prevailing opinion among CLIL teachers is that CLIL largely tallies with the content subject in the mother tongue in its content, objectives and forms of assessment. Therefore they claim to concentrate on the content of the subject and its didactics while neglecting the target language in assessment (Vollmer 2001:207). However, they often feel uncomfortable about how to judge their students in a fair way, which they very rarely admit. Rather, they deny any uncertainty and carry out holistic assessment (Vollmer 2001:228). Wolff argues that this problem "can only be solved if the approach to assessment is generally changed" (2005:8).

Because of the complex nature of CLIL, objectives for what is being measured need to be clarified. Judging content alone might be a plausible option for lessons in which CLIL is practised as it is content that is being put across to the students. Students who are taught history or geography or any other content subject, regardless of the language in which this lesson is embedded, need a pass grade and this will be a grade in content. There is no doubt that the content curriculum dictates the parameters for assessment but, as a consequence, they address only one part of CLIL. The question that still remains is if content can be separated from language at all.

This somehow narrow view is completely contradicted in the *CLIL Compendium*. There, the integration of language in assessment is highly recommended, however, with the aid of "integrated assessment tools".

> Performance Assessment of CLIL learner performance has to be sensitive to the subject-language duality inherent within many models of CLIL. Integrated pedagogical classroom learning needs to be assessed using similarly integrated assessment tools. Viewing an examination text from a solely language or subject point of view negates the trans-disciplinary characteristics of CLIL. Testing and assessment apparatus need to be introduced which allow learners to show the breadth of their knowledge and skills in relation to both content and language.[12]

Whether traditional forms of assessment comply with the requirements of the dual focus or alternative procedures have to be considered, raises an additional question. So far, as research evidence shows, merely a few suggestions towards a solution have been made (Ernst 1995, Vollmer 2001) and only one tool of assessment, which caters for the diverse demands in CLIL, has been devised (Poisel 2007).

### 3.2 Regulations in Austria and elsewhere

In Austrian schools assessment is based on the School Act of the Minister of Education and Arts (Bundesminister für Unterricht und Kunst) of June 24, 1974 about assessment in compulsory education as well as lower and higher secondary education (Leistungsbeurteilungsverordnung) BGBI. Nr. 371/1974, amended by BGBI. II Nr. 35/1997. However, there is no explicit reference to the CLIL concept in this School Act. According to the Lower Austrian School Board (Landesschulrat für Nieder Österreich) (LSI Friedl) and the Vienna School Board (Stadtschulrat) for Vienna (LSI Kschwendt-Michel), students are allowed to choose between German and the target language in exams (Nezbeda 2005:8). If it is the case that a content subject is (nearly) continuously taught in a foreign language, an extra certification is provided in the school report.

---

[12] www.clilcompendium.com

> Die Rechtsgrundlage für die Verwendung einer Fremdsprache als Arbeitssprache ist durch den §16 (3) des Schulunterrichtsgesetzes geregelt. In Bezug auf diese Grundlage besteht seit 1997 eine Verordnungsbestimmung, die es erlaubt, den Unterricht eines Gegenstandes in einer Fremdsprache auch im Jahreszeugnis zu vermerken. Damit ist die Möglichkeit gegeben, das verstärkte Engagement der Lernenden im Fremdsprachenunterricht auch ausreichend zu dokumentieren. Dies soll vor allem hinsichtlich der weiteren beruflichen Laufbahn der Absolventen von Nutzen sein. (Abuja 1999
>
> […] wenn eine lebende Fremdsprache als Unterrichtsfach gemäß §16 Abs. 3 des Schulunterrichtsgesetzes angeordnet wird: „Der Pflichtgegenstand/Die Pflichtgegenstände.....wurde/n gemäß §16 Abs. 3 des Schulunterrichtsgesetzes in der lebenden Fremdsprache ..... unterrichtet.“ Zeugnisformularverordnung § 3 Abs. 1 Z. 11a, Novelle BGBI. II Nr. 130/1997 (Nezbeda 2005:11).

While the official recommendation in forms up to the school-leaving exam (Matura) might imply that assessment in the target language is disregarded, it is given weight in the finals at an AHS. If students take an exam in a content-subject in a foreign language in the school-leaving exam (mündliche Schwerpunktprüfung), the School Act states that both content and language are assessed.

> §20 (1) Die mündlichen Schwerpunktprüfungen umfassen zusätzlich zur Kern- und Spezialfrage (§19 Abs. 3-5) [...]
>
> (2) Für die mündliche Schwerpunktprüfung hat der Prüfungskandidat [...]
>
> 3. bei der ergänzenden Schwerpunktprüfung gemäß Abs. 1 Z 3 eine Aufgabenstellung aus dem Prüfungsfach in Verbindung mit dem darauf bezogenen schulautonomen Pflicht- bzw. Wahlpflichtgegenstand (Abs. 1 Z 3 lit. a) bzw. eine Aufgabenstellung [...] in der Fremdsprache (Abs. 1 Z 3 lit. c) zu bearbeiten.
>
> Dies hat in sachlich und sprachlich richtiger Ausdrucksweise zu geschehen (Nezbeda 2005:13).[13]

These regulations allow students to take the final examination in CLIL with regard to the supplementary question (“Ergänzungsfrage”). In so doing, they have to answer

[13] http://www.bmbwk.gv.at/schulen/unterricht/reifeprüfung/vo_rp_ahs.xml#a2u1

the question in an appropriate way with regard to both content and the target language.

If students have done the supplementary question in a foreign language in the school leaving exam, it is recorded in the school leaving certification (Maturazeugnis) as follows (oral communication with T2):

e.g. Geschichte ergänzend (Englisch)

Eurydice, the statistical branch of the European Commission[14], presents an overview on assessment in European countries where CLIL education takes place. It should be noted, however, that no general, Europe-wide guidelines have been developed so far concerning how to give students credit for taking part in CLIL and how to certify their achievements.

CLIL provision is available in all countries across Europe except Denmark, Greece, Cyprus, Liechtenstein, Portugal and Iceland. In almost half of the countries offering CLIL assessment is carried out in the target language, in the others pupils use only the language of mainstream curricula.

In countries like Austria, Hungary and Ireland, pupils are offered the choice of doing exams either in the target language or the language of the mainstream curriculum. In the Netherlands, students are assessed in Dutch but may take an additional exam in English to show their language ability. In all of these countries, the linguistic 'value added' achieved by pupils is (or may be) formally recognised and certified.

Moreover, resulting from bilateral agreements between certain countries, for example, in the Czech Republic, Germany, Poland and Bulgaria, certificates in CLIL are the gateway to higher education in the respective partner countries. They enable pupils to enrol in universities abroad without having to take language tests.

[14] www.eurydice.org

Wolff (2005:20) argues that certification is still an unsolved problem; there are no general, Europe-wide guidelines so far as to how give students credit for taking part in CLIL and certify their achievements. Students' participation in CLIL is mentioned in their school leaving certificate (see above) but they do not get specific diplomas. Wolff advocates a Europe- or even world-wide agency to be established to check the standards of CLIL and deliver specific CLIL diplomas. They should allow students to undertake further education beyond their home country.

This request for unified certification is spelled out in the CLIL compendium as follows:

> Formal recognition of learner achievement in certain types of high-activity CLIL should be made at national level. Efforts to have such documentation recognised by authorities and institutions trans-nationally should follow.[15]

### 3.3 Practice of and suggestions for error correction from the literature

In traditional language classes accuracy and formal correctness are of prime importance and serve as criteria of assessment, the effects of which have proved to be discouraging and demotivating for students. In Germany, where students have the possibility to drop regular language lessons in higher secondary education, they do so in increasing numbers for this reason but continue attending CLIL classes (Ernst 1995:259). This preference for CLIL over traditional language teaching seems to be grounded in a more flexible generosity towards error correction.

> Welche bilingual unterrichtende Lehrkraft kennt nicht Schüleräußerungen wie 'Wir können im bilingualen Unterricht viel freier und lockerer sprechen, ohne immer korrigiert zu werden. Wir trauen uns, etwas zu sagen, weil wir unser Wissen loswerden wollen" (Ernst 1995:261).
>
> *Which teacher of bilingual classes does not know students utterances such as "We can speak more freely and in a more relaxed way, without being continuously corrected. We have the courage to say something because we want to demonstrate our knowledge".* My translation.

[15] http://www.clilcompendium.com/recomme.htm

CLIL is content learning in a foreign language, and the learning goal is competence in content. Assessment is based on the guidelines of the content subjects, which might imply that linguistic errors are of less importance and should be dealt with generously.

> Die Fehlertoleranz von Seiten der Lehrkraft in der Fremdsprache ist unbedingt zu erhöhen, da das Ziel einer Stunde immer fachlicher Natur sein wird, die SchülerInnen also primär Inhalte verarbeiten sollen (Abuja&Heindler 1993:17).
>
> *Teachers' tolerance towards errors in the foreign language must be increased as the goal of a lesson will always be content, the students should primarily process content.* My translation.

Nevertheless, CLIL teachers have to deal with the correction of errors in language in one way or another. The question they are always confronted with is how to handle error, both in the written form and orally.

Dalton-Puffer's in-depth discussion of error correction delineate how errors in the spoken medium are dealt with in the Austrian CLIL classroom, with respect to both factual and linguistic errors (2005:187-216). Dalton-Puffer's findings indicate that content errors, on the whole, are corrected while the same is not true for language errors. These are treated with higher tolerance as the focus in the CLIL classroom is on functional communication. Yet the degree of tolerance varies according to the type of linguistic errors. While lexical errors get most attendance from the teachers (cf. example 6, 15), errors of pronunciation receive less correction and errors of grammar are usually ignored (cf. example 7, 8). Dalton-Puffer's study provides evidence that CLIL teachers generally agree to the request for more tolerance towards errors (see above) as "for most CLIL classes language trouble and its correction appears to be a low stake issue" (2008:14).

Another interesting issue raised in this context is Dalton-Puffer's reference to the discrepancy between what teachers profess in the interviews they do and how they act in the classroom. Teachers with a qualification in EFL claim to overlook errors while those without would correct them. Actual classroom observation, however, proved the opposite. Dalton-Puffer follows that "it is rather fascinating to observe the

phenomenon (known from other contexts) that what people do and what people say they do often are two different things altogether" (2005:194, cf. chapter 6.6).

The correction of linguistic errors was already of major concern when CLIL was still in the early stages. Ernst, a practitioner and advocate of CLIL in Germany, was the first author to comment on the shortcomings of assessment in CLIL, discussing correction in detail (1995: 258-264). He argues that language errors in CLIL should not be passed over but have to be taken into account as part of assessment.

Within the linguistic dimension of CLIL Ernst (1995) suggests distinguishing between four categories of errors:

1. Errors of a phonological, morphological, syntactic, semantic or pragmatic nature that impede or impair understanding should be corrected and assessed as they would not allow effective communication. In this category he also includes the use of German vocabulary in higher secondary education and also in exams as they might irritate native speakers.

2. So-called pragmatic errors, i.e. errors of register, which are inappropriate in respect to a specific culture or situation, should be corrected. As intercultural competence is a specific learning goal of CLIL, empathy for different people and cultures should be developed. However, Ernst contends that these errors do not lead to misunderstanding and are often difficult to correct, therefore less weight should be attached to them.

3. Errors of form that occur in phonology, morphology, syntax, or lexis are deviations from grammar rules but do not impede understanding. Therefore they should be judged more leniently than in language lessons.

4. Errors in content-specific terminology, he argues, are part of content and should be corrected and assessed, in particular with regard to terms that have been previously dealt with in class.

For the correction of written exams Ernst recommends three different columns, separating errors of terminology, common language and content. His suggestion, however, seems to be unmanageable: in terms of practicality it might be too time-consuming and, additionally, it might be difficult in some cases to distinguish one type of error from the other.

It is noteworthy that Ernst takes a step in the direction of self-correction according to defined criteria. He suggests that errors, in particular of common language and terminology, should be marked by teachers but corrected by the students themselves in order to develop an awareness of the terminology and, in general, of the expectations and norms of CLIL. What he does not mention, though, are the clearly defined criteria against which the students should correct their tasks.

Vollmer (2001: 231) does address the issue of defined criteria. He contends that assessment in CLIL can only be improved if a nationally, or even Europe-wide, agreement on objective criteria at different levels of competence in content subjects can be reached. He suggests developing a framework for content subjects similar to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF). He exemplifies his suggestion with the already existing "Geography Education Standards Project " from Washington D.C. from the year 1994 (printed in Haubrich et al. 1997:29). This framework, which should contain 'can do' descriptors as in the CEF, will facilitate the definition of learning objectives and, simultaneously, lay down the criteria for assessment. Only when these criteria are established, he argues, can students turn to self-assessment. Clearly defined criteria would prove beneficial in portfolio work as well as in life-long learning.

Additionally and equally importantly, Ernst discusses the issue of specific training in assessment for both university students as well as teachers. In particular, he stresses the need to gain experience in practical work. He argues that students should not only study theories of assessment but also, first and foremost, become familiar with the practice of assessment in order to develop and show their competence and also raise their awareness of linguistic errors. For teachers, he demands broadening their experience of assessment as well as discussing their norms of assessment at workshops or meetings as "es ist ein unhaltbarer Zustand, daß eine Rexflexion der

eigenen Normen kaum erfolgt"*(It is intolerable that teachers hardly ever reflect their own norms) (*My translation).

Ernst was not only the first to comment on the correction of linguistic errors in CLIL and their assessment. For a long period of time he was also the only one. His suggestions seemed not to have been welcomed by teachers; at least in public they did not trigger off any reactions. Thus, it might be assumed that every teacher develops their own modes and tools of assessment to do their students justice. The only chance, however, to progress satisfactorily in this field would be an open and intense discussion among the teachers involved in the CLIL process (Vollmer 2001:208).

Oral communication with students and teachers revealed that by the year 2008, things do not seem to have progressed much further. Neither have students in the final stages of their academic teacher education, been introduced to questions of assessment at university, no matter whether in general or CLIL- specific, nor do teachers discuss and exchange their experiences with colleagues. However, with the issue of assessment gaining in importance there are clear signs of change. University courses are now being offered for students; they not only discuss theories of assessment but also require participating students to experiment with and reflect on possible methods for assessing their own achievements, foregrounding the process of learning.

### 3.4 Practice of and suggestion for assessment from the literature

As mentioned before, research literature on assessment in CLIL is scarce; if the issue is addressed, the problems are raised and/or identified and demands for solution put forward. Fehling and Finkbeiner (2002:22-32) went one step further. They carried out a survey in order to evaluate the effectiveness of assessment practices in CLIL and discussed and reflected on their findings.

At a CLIL training workshop, 25 CLIL teachers were asked how they assess written exams. The questions mainly concerned the weighting of content and language in

exams, the correction of language errors and possible legal regulation and/or school-internal guidelines, the teachers could follow.

The teachers' answers show a high degree of consensus on the one hand, but diverge radically on the other. They generally agree that they focus on the content of the subject. The answers as to whether language errors are assessed, however, significantly differ. About half the teachers correct language errors but do not judge them, the others also assess them. Some of the latter additionally distinguish between assessing all linguistic errors, particularly if they impede understanding, and merely errors in the terminology of the respective subject. The differences, Finkbeiner argues, can be explained with the absence of internal and external guidelines that a teacher can adhere to.

The findings of the survey identify the fundamental shortcomings of the assessment method applied. There is a glaring discrepancy between what is assessed in the exams and the goals of CLIL, such as the enhancement of language competence as well as language and cultural awareness; none of these aspects is taken account of in the written exams. Finkbeiner emphasises that unfortunately no assessment tools exist that cater for these aspects.

With regard to set criteria, Finkbeiner draws on Mäsch's demand that

> die Bewertung der Schülerleistung im Fremdsprachenfach, z.B. im Fach Französisch, soll außenrelevant und nicht binnenrelevant erfolgen. Denn für (absolut) 'hervorragende' Leistungen muss es ‚hervorragende' Noten geben (Mäsch 1991:51, quoted in Finkbeiner.2002:28).
>
> *//the assessment of students' achievements in foreign language subjects, e.g. French, should be judged against external and not internal criteria. Because for (absolutely) "excellent" achievements students should be awarded "excellent" grades//*

Her suggestion, however, does not imply that the prevailing internal criteria against which students are judged should exclusively be replaced by external criteria, rather she proposes a combination of both. She argues that blending these criteria would allow for both fair and valid assessment at school and according to national and international standards.

If CLIL is to advance successfully, she contends, research in the area of assessment should be intensified, criteria need to be unified, and teachers should be trained how to assess students' achievements in relation to the various goals of CLIL.

The need for change has been recognised (Fehling and Finkbeiner 2002:22-32, Wolff 2005:8), but only one suggestion for new instruments has been put into practice. Poisel (2007:43-46) presents the model *'Assessment Modes in CLIL to Enhance Language Proficiency and Interpersonal Skills'*, which "involves process writing and formative assessment, combined with portfolio work and peer tutoring" (2007:43). This model is based on portfolio work, which students carry out individually with support from their teacher and their peers. It meets the needs of students' language enhancement, acquisition of content knowledge, communicative competence and interpersonal skills.

The portfolio work is organised in 4 stages. It begins with the teacher instructing the students what they have to do for their portfolio work within a definite time frame and how to present their findings. Special emphasis is placed on the language element by providing students with a list of *I/you can do…*descriptors from the CEF for all skills; these enable the students to do self-assessment and peer tutoring. Additionally, the students are taught how to give constructive and encouraging feedback to their peers. The second step has to do with the actual work process. It encompasses the collection of information on the content topic and its critical selection. Students learn how to come to terms with unknown words and develop reading strategies. They then write texts such as analysing or commenting on their findings. These texts are shown to their peers, who are expected to give descriptive feedback. After the modification of the texts according to their peers' suggestions, the texts are presented to the teacher, who reads them and gives their feedback. It is then left to the students to take up the teacher's advice or follow their own design. The third step concerns the students' presentation of their final work. In the fourth step the teacher finally assesses the student's work, focussing rather on the progress the students have made in the course of their work than on the outcome, and on the quality of their presentation.

It needs to be mentioned that the portfolio work suggested by Poisel is designed for an international school, in which the language objectives for each level are defined by the CEF. Furthermore, due to the international character of this school, it can be assumed that the language of communication among peers is English. The situation in Austrian mainstream schools is different: a curriculum for the Austrian CLIL classroom, which establishes linguistic criteria, has not been developed so far. And the language Austrian students use among each other at school is nearly exclusively German.

# 4 The role of language as a medium of learning

Since assessment cannot be examined in isolation from other aspects of the educational process, I will briefly discuss how knowledge is constructed in the CLIL classroom. While it is true that in CLIL learning takes place using a medium other than the students' first language, language also plays a key role in the acquisition of knowledge in general. Thus I think it is important to explain the wider implications for teachers and students of teaching and learning content in a foreign language. How teacher and students do it depends on the methodology that is applied in the CLIL classroom. Hence I will view the methodology suggested for and actual classroom practice in this innovative approach.

## 4.1 Language across the curriculum – the paradox

The notion of language across the curriculum refers to the fact that language learning does not only occur in language subjects, such as mother tongue education or foreign language learning, but also occurs in every subject at school across the whole curriculum. Consequently, every subject teacher is also a language teacher. This indicates that not only language subjects cater for the development of language skills and abilities but emphasises that language is central to the whole curriculum process, i.e. language development lies within the responsibility of all subjects of the curriculum (Vollmer 2006). The linguistic dimension that is responsible for successful learning in every subject has been fully acknowledged in English-

speaking countries. In the German-speaking world, however, the awareness of language as a medium of learning that has to be fostered does not seem to have permeated into all subjects of the curriculum. If content subjects are taught in the mother tongue, there is no awareness among teachers that language teaching takes place simultaneously. If, however, a content subject is taught in a foreign language, this process becomes conscious and the language is made "visible" (Dalton-Puffer 2005:9).

In CLIL, priority is given to satisfying the demands of the curriculum of the content subject. Yet language is needed to transmit the concepts of the content and to enable the exchange of thoughts and ideas between teachers and students and among students. No matter which skill they apply, listening, speaking, reading or writing, they do it through language. Mohan argues that

> language is a system which relates what is being talked about (content) and the means used to talk about it (expression). Linguistic content is inseparable from linguistic expression.
>
> In subject matter learning we overlook the role of language as a medium of learning and in language learning we overlook the fact that content is being communicated (1986:56).

Vollmer (2001:230) shares this view and contends that language, being the main means of learning at school, should become a major issue in the educational culture. The role it plays for learning in general, and for learning in CLIL in particular, has to be clearly and fully defined. He argues that:

- Jedes Lernen ist sprachliches Lernen. (*Every learning is language learning.* My translation).
- Jede intellektuelle Leistung ist sprachlich vermittelt. *(Every intellectual achievement is packaged in language.* My translation).

In view of this, the separation between content and language in assessment seems to be impracticable as language is the primary evidence that teachers have for judging students' achievement. Students demonstrate their factual knowledge by means of language as a resource for understanding. If, for instance, communication breaks

down, how then could students possibly display the knowledge they have of the content?

### 4.2 The construction of knowledge

While theoretically CLIL would suggest the teaching of both content and language, research literature reveals that there is not much likelihood that language is explicitly taught:

> We SPEAK English in our GEOGRAPHY class, and we LEARN English in our ENGLISH class (Gröne quoted in Thürmann 2000:80).

Therefore it can be assumed that the language used in the classroom is solely responsible for the language competences CLIL students might achieve. It is commonly acknowledged that other than in traditional language teaching, CLIL is supposed to bring real-life situations into the classroom, which allows for meaningful communication. How then is this parcel of knowledge packed in such a way as to lead to meaningful communication? How is subject matter structured and finds its way from one mind to the other in the classroom? Two strategies seem to be predominantly used in CLIL classrooms in Austria: the IRF pattern and questions.

**The IRF pattern**

CLIL classes, like all content-classes are predominantly oral events. How students get to use language is profoundly influenced by the activity types that occur in the classroom. In the Austrian CLIL classroom, the most common practice of teaching is whole-class interaction. It "consists of the teacher conducting a dialogue with the class as a collective conversational partner" (Dalton-Puffer 2005:39.) Teaching is led by the teacher, who calls for volunteers or nominates students. The typical conversational interaction between teacher and student consists of a three-part structure: Initiation (I), Response (R), and Follow-UP (F). This structure has come to be known as the IRF pattern (Dalton-Puffer 2005: 63 ff).

The IRF structure represents the

> very foundation of language-based learning: The Initiation opens up that space where the already existing cognitive structures of the learner become accessible to further development, expansion and change, that is learning (Dalton-Puffer 2005:67).

While the Initiation move is usually controlled by the teacher, used to introduce new information, the Follow-Up move serves, on the one hand, to validate students' correct response and, on the other, more importantly, it is the move where guided construction of knowledge takes place. Generally, information is very rarely transferred from one mind to the other without disturbance. It is rather negotiated and co-constructed in a longer process between teacher and students in collaboration (Dalton-Puffer 2005:64).

Here are two examples, taken from my data, for illustration:

IRF pattern:
example 1:

I 1 T: Who followed Batista in Cuba?
R 2 S: Fidel Castro.
F 3 T: Exactly, good

Triadic dialogue :
example 2:

I 1 T: What was the situation like in Austria after the end of WWII?
R 2 S1: no food
F+I 3 T ok. anything else?
R 4 S2 many houses were destroyed
F- 5 T Have you ever heard of 'the four in the jeep'?
R 6 S3 die besatzung, eh, the aliates.
F 7 T the allies, yes, the four allies.
R 8 S4 they were in the first district
F+I 9 T yes, good, and where else?…..

The Initiation is normally a question which introduces a new topic or revises old material. The response is the student's contribution, which is then evaluated as fitting or not. This evaluation, however, need not necessarily occur and the student's answer is taken as a basis for further development of the topic. The IRF structure has become "a stable constitutive element of classroom interaction" (Dalton-Puffer 2005:65) because it is "one specific element of communicative behaviour at which culture is transmitted (Stubbs 1983:123, quoted in Dalton-Puffer 2005:65).

**Questions**

In a CLIL lesson in which teacher-led instruction dominates, the development of knowledge is promoted by strategic questioning by the teacher. Learning matter is formulated by the teacher in the target language and presented to the students. In such an environment, discourse is reduced to answering questions on the information presented. The CLIL classroom thus presents itself as cognitively undemanding: questions addressing facts heavily outnumber those for explanations, reasons or opinions (Dalton-Puffer 2005:108). Dalton-Puffer states two reasons that explain the overwhelming dominance of questions about facts, one referring to the time-pressure teachers' feel to cover the content of the curriculum (2005:108), and the other to do with the teachers' language competence and/or confidence; teachers feel safer if the discourse in the classroom remains within the topic area they have planned for the lesson and is thus predictable (2005:107).

Questions asking for facts are generally responded to with extremely short phrases, often consisting of only one noun, a verb or an adjective (Dalton-Puffer 2005:96). Questions that produce more elaborate output are cognitively more demanding but do not occur frequently.

**Academic language functions**

Academic language functions are a particular set of communicative functions of language, which occur in an academic context and are deemed to foster critical thinking. They are not bound to a certain subject but can be transferred to other disciplines (Dalton-Puffer 2005:109-110).

The following list comprises the major academic language functions (Dalton-Puffer 2005: 110):

Analysing
Classifying
Comparing
Defining
Describing
Drawing conclusions
Evaluating & assessing
Hypothesising

Informing
Narrating
Persuading
Predicting
Requesting/giving information

Dalton-Puffer examined three typical academic language functions in the CLIL classroom, namely defining, explaining and hypothesizing (2005:109 ff).

*Definitions* are hardly ever found in the Austrian CLIL classroom, which cannot possibly be ascribed to students' inability to master the structure, as it is very basic. Rather it seems to be the case that students are not familiar with definitions. Thus, to clarify meaning of unknown terms, only teachers, though very rarely, make use of definitions. In general, both teacher and students prefer translating the unknown term into German.

*Explanations* occur in the CLIL classroom on a regular basis but again in low numbers. The answer might again lie in the fact that there is no awareness of this speech function in the classroom.

*Hypothesising* is even less frequent than defining and explaining. Here, the absence may be grounded in the complexity of grammar required for this function. Another reason might be that the discourse in the classroom is mainly about facts and very rarely about reasons, beliefs and opinions, which tends to exclude hypothesizing.

My own observation of oral exams in CLIL classes shows that definitions indeed never occur. However, explanations and hypothesising are frequently found in my data, as the teacher (T2) not only asks content questions that have been dealt with in the previous lessons, but adds thought-provoking question such as "can you explain?", "what does it mean ?", "what do you personally think?" on a regular basis (see examples 11, 12, 13). While explaining was rather easily handled by the students, hypothesising proved to be extremely difficult, in particular when the student answered the teacher's question "what might have been if…" with a conditional clause (see example 8).

## 4.3 Teaching methodologies

> One difficulty in thinking about knowledge is that it is both "out there" in the world and "in here" in ourselves. The fact that it is "out there" and known to a teacher does not mean that he can give it to children merely by telling them. Getting the knowledge from "out there" to "in here" is something for the child himself to do – the art of teaching is knowing how to help him to do it. (Barnes 1976:79, quoted in Wolff 2002:1)

In the CLIL discussion the call for new assessment tools (cf. chapter 3.1) goes hand in hand with introducing new teaching methodologies. While assessment, however, is still fraught with problems, a CLIL-specific methodology is already being discussed in concrete terms. It suggests meeting the specific demands of the integration of subject learning and language learning or, as Wolff words it, "an approach which is content-oriented but at the same time language-sensitive" (2005:17). The key question is how such complex learning processes can be promoted methodologically to do best justice to this demand.

Immersion programmes as practised in Canada, Australia or America are deemed to be inappropriate for direct transfer. They work in countries where students have direct and regular contact with the foreign language outside school; that is to say they have more exposure to the target language in the community and through the media. These programmes cannot simply be transferred to a predominantly monolingual country such as Germany, and likewise Austria, in which students' exposure to the foreign language is almost entirely restricted to school. In such countries the need for a methodology specially tailored for CLIL, which integrates subject and language work, is of central importance (Otten 1993:50, quoted in Thürmann 2000:78).

Nevertheless, there is no consensus among CLIL advocates. Opinions regarding this issue differ and are expressed in two opposing schools of thought. One firmly denies a specific methodology, for example Nando Mäsch (1996), who does not approve of a specific methodology for CLIL. He believes that there is no necessity for pedagogical innovations if one adheres to two established principles that underlie good educational practice: the principle of utility and the principle of generality.

> Erlaubt ist, was der Sache dient und dem Schüler nicht schadet; Leitsatz der Utilität nach Kronenberg (1993) – wobei mit „Sache" der bilinguale

> Fachunterricht gemeint ist. (*Everything is allowed that serves the matter and does not do any harm to the student; principle of utility by Kronenberg (1993) – whereby "the matter" refers to bilingual content teaching.* My translation)
>
> Für das bilinguale Lehren und Lernen gibt es keine anderen Unterrichtsformen als die heute für guten schulischen Unterricht bekannten; Leitsatz der Generaliät (Mäsch 1996:1f, quoted in Thürmann 2000:77). *For bilingual teaching and learning there are no different methodologies than those known in good school teaching today.* My translation)

In view of the dual focus on content and language in CLIL he admits, though, that particular aspects have to be considered; yet he refrains from specifying these characteristics in detail (Thürmann 2000:77-78).

Others, like Thürmann (2000), Wildhage and Otten (2003) are more progressive and defend the need for a specific methodology in CLIL education. Thürmann emphasises that content learning is and must remain at the core of the lesson, allotting it as much time as possible (2000:80). Wolff argues that there should be no difference in the subject knowledge CLIL students gain in comparison to their peers who study the content subject in their mother tongue (2005:6). However, students acquire the content of the subject through a language still being learned. In other words, both content and language are explored in the CLIL lesson. In order to cope with the content in the foreign language students need to be equipped with linguistic knowledge and skills, which enables them to fill the gap between what they might want to say (content competence) and what they are able to say in the foreign language (language competence) (Bach 2000:19).

This gap can be bridged fairly easily as long as CLIL is taught in the traditional way, in which teacher-centred instruction governs the classroom. Teaching material can be adapted and tailored with regard to both content and language, reducing language teaching and learning to content-specific terminology. The only "challenge" students face is answering questions about the texts provided. If however, student-centred activities for the construction of knowledge find their way into the CLIL classroom, this gap becomes remarkably wider (Thürmann 2000:76). While in teacher-centred instruction students' language production is predictable, in student-centred construction of knowledge the gap between content competence and language competence needs to be bridged.

Responses to this dilemma are, next to organisational and didactic suggestions (see Thürmann 2000:81), the development of a specific methodology "to expand the students' interlanguage in order to become compatible with the content objectives and working methods" (Thürmann 2000:82). However, language work should only be integrated if the need arises (Wolff 1996).

Adherents of a methodology tailored for CLIL place special emphasis on the following language-oriented issues; these are not subject-specific but extend across the disciplines (Wolff 2005:10):

- receptive processing
- language production
- autonomy

*Receptive processing*

Thürmann draws on Mohan's distinction between developmental reading (learning to read) and functional reading (reading to learn) (Mohan 1986:14, quoted in Thürmann 2000:85) and suggests the implementation of the latter in CLIL. As students come across different text types in CLIL, they need to master specific reading skills to understand the concepts in them and gain the information they contain (skimming, scanning). Receptive processing not only refers to reading texts, students also should learn how to infer the information depicted in visual aids, such as graphs, maps, charts, pictures or cartoons.

*Language production*

"Content-based instruction is particularly effective when it combines a focus on content with a focus on form" (Mohan and Beckett 2003, quoted in Dalton-Puffer 2005:250). In terms of methodology we need to ask then how to integrate content and language work. CLIL teaching is primarily subject teaching, which means that the subject offers the content the students have to work with. Researchers and practitioners have agreed that the discourse necessary for dealing with the content in the CLIL classroom consists of functional skills.

Phrases for classroom discourse were developed almost simultaneously by Abuja in Austria and Otten and Thürmann in Germany. They conclude that there are basic learning and communication strategies, which are transferable from one subject to the other. Abuja presents functional categories such as Definition – Classification – Describing states and processes – Working with graphs – diagrams, tables, etc... Otten and Thürmann resort to academic functions such as Identify-Classify/ Define-Describe- Explain/Argue-Evaluate-Reduce-Elaborate-Exemplify-etc. CLIL, taking place in academic setting, can be expected to support the development of these functions (Thürmann 2001:88).

Dalton-Puffer's investigation underpins the need to introduce academic language functions. It also reveals that the reality of the Austrian CLIL classroom falls short of these expectations: active use of academic functions in classroom talk is almost entirely absent (2005:109ff).

In order to counterbalance the prevailing Triadic Dialogue that leads to rather minimal output on the part of both teacher and students, Dalton-Puffer recommends the re-introduction of the teacher monologue "(in well-considered dosage) both in the interest of presenting coherent conceptual networks of topic content and in the interest of providing sustained, syntactically complex oral input" (2005:249-250).

Additional aspects concerning changes in how to deal with the language used in CLIL need to be mentioned. Monolingualism is no longer an issue; the subject content should not exclusively be taught in the foreign language. If students are expected to develop cultural awareness, they need to compare and contrast material in both the foreign language and the mother tongue (Wolff 2005:18). The same applies to code-switching. While at the beginning of CLIL the mother tongue was not to be used in the classroom (in conformity with the theory of language teaching at that time), code-switching is now seen as an important element for raising language awareness (Wolff 2007:3).

The teaching of content-specific terminology, which was at the core of CLIL in its early days, has lost importance. It is suggested that students are provided with more general subject-oriented content terminology and then progress to more specific vocabulary (Wolff 2005:17).

*Autonomy*

CLIL is assumed to offer the appropriate learning environment in which to integrate concepts developed by modern learning theories such as constructivism and co-operative learning. According to constructivist theories, students do better when they are not passive recipients of knowledge but are actively involved in the learning process and take on responsibilty for their own learning. If they work actively with one another in partner- or group-work, they tackle the subject content more consciously and more emotionally (Wolff 2007:1).

The suggested methodology meets the dual-focussed objectives of CLIL by supporting both the learning of content and language. It is solid and convincing. Yet, it remains to be asked whether it has been translated into classroom practice. CLIL, being an innovative approach, could have promoted pedagogical innovation, as was voiced by Nikula at the CLIL Symposium in Vienna, 2007: "I said already ten years ago that CLIL could have been a catalyst" (oral communication). This opportunity, however, seems to have been embraced with little enthusiasm:

> „Most schools seem to be of the opinion that simply the choice of CLIL is innovative enough, and that it is not necessary to expect any further pedagogical innovations through the implementation of CLIL (Wolf 2005:5).

Wolff is referring to the current situation of innovative methodologies employed - or rather not employed - in CLIL classes in Germany. This situation, however, can be encountered in Austria, too. Dalton-Puffer's extensive investigation (2005) of the CLIL situation in Austria and my own experience both exhibit that the way content subjects are taught in a foreign language is no less traditional in our country.

# 5 Empirical Study

The previous chapters have shown that there is a strong desideratum for appropriate procedures and tools of assessment in CLIL while no empirical research on assessment practices in CLIL exists so far. Considerations and tendencies steer towards a deviation from the traditional modes in favour of alternative ones. Whether teachers in the Austrian CLIL classroom have taken up these suggestions is difficult

to ascertain as classroom teachers tend to be reluctant to show their assessment practices (see below).

## 5.1 Access to the field

The aim of this study is, therefore, to find out how assessment is carried out in practice. For this purpose it is necessary to investigate actual lessons, which I intended to do by video-taping. This, however, proved to be a major obstacle. It took me several attempts to find willing teachers until eventually two colleagues opened the doors of their classrooms. In fact, there was only one teacher (T2) who readily agreed to participate in my research project. With some difficulty, she succeeded in talking a colleague of hers (T1) into going along with my plan. This endeavour reinforces the fact that assessment is a very private affair and teachers seem not to want to have anybody look over their shoulders (cf. chapter 7.2).

However, it was not only the teachers who made the project difficult. When I had obtained the teachers' approval and the permission of the school inspectorate (see appendix), it was the students who revolted and refused to be video-taped while being assessed. It cost T2 considerable effort to help me get my project going. She explained it (and its potential benefits) to one student, who she assumed to be most influential in class, and asked him to persuade his peers to participate. Her plan worked well so that eventually all the students but one allowed video-taping.

Nevertheless, one teacher (T2) even welcomed my research:

> Ich bin froh, dass sich jemand mit dem Thema befasst und dass es untersucht wird. Wir arbeiten hier ja in einem Vakuum, keiner weiß etwas vom anderen (T 2 at my first encounter at school).
>
> *//I'm glad that someone is addressing and investigating this topic. We're working here in a vacuum; nobody knows anything about the other one//*

My initial intention was to investigate assessment in a form of lower- and upper-secondary education, respectively. However, the school to which I was allowed access offers CLIL only in upper-secondary forms. There I then investigated assessment in a fifth and eighth form.

## 5.2 Research design

The empirical part of the present study is a qualitative study comprising classroom observation and videotaping, interviews with teachers, comments on transcribed exams, documentation, protocols and personal notes (cf. Flick, Kardoff, Steinke 2000: 246).

Since assessment in CLIL is uncharted territory, a qualitative research methodology was the obvious choice because:

> Qualitative Forschung ist immer dort zu empfehlen, wo es um die Erschließung eines bislang wenig erforschten Wirklichkeitsbereichs („Felderkundung“) mit Hilfe von “sensibilisierenden Konzepten“ (Blumer 1973) geht. Durch den Einsatz von „naturalistischen Methoden“ wie teilnehmender Beobachtung, offenen Interviews [...] lassen sich erste Informationen zur Hypothesenformulierung für anschließende, standardisierte und repräsentative Erhebungen gewinnen (Flick, Kardoff, Steinke 2000:25).

In addition to classroom observation and video-taping of the exams I carried out interviews with the teachers, concentrating on their practice of assessment. I decided on focussed interviews because they centre on a predefined topic and attempt to trigger off reactions and interpretations in a relatively open way. Focussed interviews were originally group interviews, which are similar to structured interviews, but offer the interviewer more flexibility with regard to the interview guideline and possible responses on the part of both the interviewer and the interviewee (Flick, Kardoff, Steinke 2000:353).

> […] die im Rahmen teilnehmender Beobachtung durchgeführt werden, in denen spezifisch gemeinsam erlebte Situationen – z.B. Unterrichtssituationen – abgehandelt werden (Flick, Kardoff, Steinke 2000:354)

The quality of focussed interviews is determined by four criteria. These are:

1. Range: The interviewee should be granted the maximum of freedom and flexibility to react to the stimulus situations.
2. Specifity: Topics and questions in the interview should be answered specifically and not globally.

3. Depth: The interview should be representative in relation to affective and cognitive importance that certain situations have on the interviewee.
4. Personal Context: The personal context should provide additional information for interpretation.

(Flick, Kardoff, Steinke 2000:355-356).

The interview guideline developed for this study covered the following areas of interest:

- format and frequency of assessment
- choice of language
- weighting of content and language in grades
- anything else that is of interest

The teachers were also asked about their qualifications and how long they had been teaching in general, and CLIL specifically (cf. guideline in appendix).

Classroom observation, video-taping, carrying out interviews and eliciting comments on transcripts of oral exams represent the core of data. In addition, I made field notes and collected teaching material used in T2's lessons. The field notes were specifically used for asking the teacher after the lessons which grades she had given the students for their performances. The teaching material provided information on the topics dealt with in the lessons.

### 5.3 The institutions

Both schools from which I collected my data are grammar schools. The target language used in CLIL in either school is exclusively English. The grammar school whose teachers gave me the interviews have CLIL in their curriculum from the 1st up to the 8th form in at least one subject; in some forms two or more subjects are taught in English. The second school, which is the one where I carried out my observation, offers CLIL only in the upper school, that is from the 5th to the 8th form, usually in one subject. All the data collected and analysed are derived from history lessons taught in English, respectively from teachers teaching history in English.

At this point I would like to express my deep gratitude to the teacher (T2) who allowed me to observe her assessment procedures. Without her willingness and active support I would not have been able to write this paper. Although she seriously doubted whether my investigation would yield any research results, expecting no co-operation from her colleagues and students, it worked. How it worked and in particular how well CLIL worked so that about half the students of T2's form went in for the school-leaving exam in the CLIL subject at the end of the term will be briefly analysed in chapter 7.1.

## 5.4 Description of data

The database for this project contains the video-tapes of 11 lessons from the upper-secondary level of a grammar school, 3 lessons from the fifth form (aged 15) and 8 lessons from the eighth form (aged 18), respectively.

In order to gain more information on assessment practice I carried out four interviews with CLIL teachers: two of them with the teachers whose lessons I was allowed to observe and, additionally, two interviews with teachers from a different grammar school. The total length of the video-tapes is 550 minutes, or nearly 7 hours; the total length of the recording of the interviews is 183 minutes, or 3 hours. T1, T3 and T4 were interviewed for approximately 20 minutes each. For reasons that will be discussed later, the interview with T2 took 114 minutes (see chapter 6.5).

The following table provides a detailed overview of the data used for the analysis in this study:

**Video-tape**

| | | | |
|---|---|---|---|
| T1 | fifth form | 3 lessons | 150 minutes |
| T2 | eighth form | 8 lessons | 400 minutes |

**Interviews**

| | | | |
|---|---|---|---|
| T1 | history in English | fifth form | 22 minutes |
| T2 | history in English | eighth form | 114 minutes |
| T3 | history in English | fourth form | 21 minutes |
| T4 | history in English | fourth and eighth form | 26 minutes |

The recordings were made between December 2007 and March 2008.

The video- and audio recording equipment was always visible to the participants, which caused some kind of uneasiness among both the teacher and the students at the beginning of my observation. After two lessons or so, however, they stopped paying attention to it. The recorded material was then transferred to the computer for transcription purposes.

The recordings of the exams were transcribed using a simplified version of the conversation analytic conventions put forward by Atkinson and Heritage 1984: ix-xvi, adapted by Markee (2002) for the purpose of studying classroom interaction (see appendix).

9 transcribed oral exams were commented by a teacher (T2) and were used for further clarification of the role the language plays in assessment.

As mentioned above, two teachers allowed observation of their classes. One teacher (T1) carried out assessment in the last two or three lessons of the term, which decided on the final grade the student had obtained. The other teacher (T2) assessed her students throughout the term, which gave me the opportunity to observe whole lessons.

In the classes I observed, English was used as the medium of instruction but was not consistently spoken throughout the lesson. For directives the teachers sometimes, very rarely though, switched from English to German. As I see it, the purpose was to make themselves better understood. What I could observe in the classroom and from listening to the recordings is that English was not used as the language of

communication in peer interactions; students preferred to speak German when talking to each other.

# 6 Assessment in the Austrian CLIL classroom

## 6.1 The teachers and their attitudes towards CLIL

The teachers in this study hold a dual qualification as language teachers of English and subject teachers of history. They have extensive teaching experience in both the subject matter and in English, and have been teaching CLIL for eight to ten years. None of them received special instruction for the teaching of a content subject in the foreign language during their pre-service or in subsequent in-service training.

> T2: Da gab es nie ein Angebot. Ich war zwei- oder dreimal bei inoffiziellen Treffen, die eher ein Erfahrungsaustausch waren und von Kolleginnen initiiert.
>
> //Nothing was offered. I've been at inofficial meetings two or three times, which were rather an exchange of experiences, initiated by colleagues//

Neither did any of them receive training in assessment as part of their university courses. (see chapter 3.1). T2 said that she had her first experience of assessment in her first year at school and adopted the modes of assessment used by her mentor (Einführender), and then developed her own way of assessing.

> T2: ich habe mir das von meinem Probelehrer abgeschaut. Und dann habe ich halt meinen Stil entwickelt.
>
> *//I copied it from my mentor. And then I developed my own style//*

The teachers' attitudes towards CLIL differ markedly. While T1 is a committed history teacher and deeply regrets the loss of content in history, T2 is not so much concerned about less coverage. She favours history in English because:

> Ja, ich finde es eine tolle Sache, weil realistisch gesehen, glaube ich, dass so oder so aus meiner Erfahrung als Schülerin, das was ich mir vom Geschichtsunterricht gemerkt habe, waren 3, oder 10%. Ich glaube, dass es eine Illusion ist, dass sich die Schüler Details merken. Um die Sprache zu forcieren, glaube ich, ist es wichtiger, einfach mehr zu reden als herunterzubeten auf Deutsch, z.B. die Habsburger Thronfolge.

//Yes, I think it's great because, realistically, from my experience as a grammar school student, what I remember from history were 3 or 10%. I think it's an illusion that students can remember details. To push language, I think, it's more important to speak than to memorize in German, for instance, the Hapsburg line of succession//

T3 claims that in terms of language the students profit from learning content subjects in English:

Es ist keine Hürde, es ist sicher ein Gewinn, wenn man das auf Englisch macht, noch in einem Gegenstand Gelegenheit haben, ein bisschen mehr zu sprechen, sich ein bisschen mehr zu beschäftigen. Diese Klasse, ich glaube die haben auch Geografie auf Englisch. Ich glaube fast in Zeichnen auch.

//It's not an obstacle, it's surely very useful if one does it in English, to have the chance to speak a bit more, just to do a bit more. This form, I think, also has geography in English. I think also art//

T4 thinks that the loss of content is balanced by the students' improvement in English as a result of more exposure to the language:

Ja, an sich finde ich es sehr gut. Ich finde es sehr günstig für die Sprache. Da werden sie schon besser. Und das zweite allerdings, das Geschichteherz blutet zwar ein bisschen, weil man natürlich vom Stofflichen sehr, sehr viele Abstriche machen muss. Ich glaube für jemanden, der sich sehr für Geschichte interessiert, ist es sehr schade, dass es auf diese Art und Weise funktioniert. Und für den anderen Schüler, ich meine, wenn wir so reden, sie lernen für die Prüfung und vergessen es wieder, oder es kommt in eine andere Lade, und da denke ich mir okay, da haben sie dann wenigstens den sprachlichen Zuwachs.

//Yes, on the one hand I think it's good. It's beneficial for the students' language. They do improve. On the other hand the "history heart" is bleeding a bit, as you obviously have to leave out a good deal. I think for someone who is very much interested in history it's a pity that it works this way. And for the other students, I think, if we talk they learn for the exam but forget it again, or it's shifted into a different pigeon hole, and so I think it's o.k., they've at least acquired more language//

T2 strongly advocates the implementation of CLIL but has experienced the students' fear of this unknown "adventure", which more often than not comes as an unexpected shock to them. They are afraid that their linguistic proficiency in the foreign language is not sufficient to follow the lessons. In order to soften this unpleasant experience in a pedagogical way and to make her students familiar with content teaching in a foreign language, T2 embeds thematic units in the regular English lessons in the 4th form. They serve as a bridge between the traditional kind of language and subject teaching in isolation and the innovative approach of teaching language and content in integration. The learning aims of these modules are slightly different from regular CLIL lessons. They are implemented to dispel students' fears and anxieties and, moreover, they are meant to make students aware of how important a foreign language can be to understand a content subject. In her "introductory modules" T2 focuses on topics which lend themselves to be taught in the English language, such as 'The Industrial Revolution', 'Victorian Women' or 'Native Americans'. These topics are actually elements of the English curriculum but in fact differ little from those covered in content lessons.

T2 encourages her students to discuss the topics and express their own opinions and beliefs. In the "module" I was allowed to observe the students compared and contrasted the position of women in the 19th century with the 21st century. Their active participation in discussions, voicing their opinions and beliefs, is of major importance for the teacher and influences the final grade to a great extent.

In terms of assessment it might be interesting to find out if the teacher makes a distinction between grading in the "modules", which are part of English lessons, and in the CLIL lessons, where the focus is exclusively on the content. A comparison between these two different subjects might contribute to clarifying the role the language plays in CLIL. This, however, is beyond the scope of this paper.

## 6.2. Components of assessment

### Participation in class

Participation in class normally earns the students a "plus" for useful contributions.

T2, T3 and T4 attach major importance to students' participation in class; T3 even ranks it as the most significant factor contributing to the grade the students can achieve:

> T3: prinzipiell zur Note trägt vor Allem die Mitarbeit bei, das ist ganz klar, wie sie sich in der Stunde verhalten, wie sie mittun.
>
> *//In principle, participation in class contributes to the grade, that's clear, how they react in the lesson, how they contribute//*

Participation in class is also highly valued by T2. She rewards her students with a „plus" when they make contributions to topics under discussion as well as to other students' exams, that is when they "help out" their classmates if these get stuck.

> Eigenständiges Einmischen in den Unterricht ist für mich sehr wichtig, oder Vergleichen, wie zum Beispiel das heute ist mit den billigen Rohmaterialien, und das wird billig hergestellt und dann teuer verkauft. Und wenn dann einer aufzeigt und sagt, ja, das ist im Grunde genommen auch Sklaverei, da schreib ich mir dann ein „Plus" auf. Mitarbeit ist wenn ich irgend eine Frage stelle und so was denkt ihr oder habt gehört, bei brainstorming auch **Noch dazu, wenn es in Englisch gesagt wird** *(my emphasis*)**.**
>
> Mitarbeit ist auch sich bei einer anderen Prüfung einzumischen oder vielleicht auszuhelfen, also wenn jetzt jemand hängt und zeigt auf und würde sagen wie z.B. „loss of identity" wie Sx.
>
> *//Independent participation in the lessons is very important for me, or making comparisons, for instance, how the way it is today with cheap raw materials, things are produced cheaply and sold at vast expenses. And then someone puts their hand up and says that that's basically also slavery, then I write down a "plus". Participation is if I ask a question like what do you think or have heard, with brainstorming, too.* ***Especially, if it comes out in English****//*
>
> *Participation is also to contribute to another student's exam or perhaps help if someone doesn't know the answer and puts their hand up and might say, for instance, "loss of identity" as Sx//*

T2 also points out that, opposed to some students' belief, reading a text aloud in class is not participation.

> Mitarbeit ist bei mir nicht Lesen. Was viele Schüler oft glauben, muss ich oft immer wieder sagen, es ist nicht aufzuzeigen und ein Stückchen von einem Text zu lesen. Das ist einfach für mich eine Selbstverständlichkeit, das wird in keinsterweise zur Mitarbeit gezählt.
>
> *//Participation for me is not reading. What many students often think, I often have to repeat it, is not putting up their hand and reading part of the text. That for me is a matter of course, and is in no way counted as participation//*

**Oral exams**

T1 believes in oral exams in principle, as she wants her students to develop oral skills.

> Sie sollen reden lernen, mündlich soviel wie möglich machen können. Nein, nein, nicht schriftlich, nie.
>
> *//They should learn speaking, be able to do as much orally as possible. No, no, not in written form, never//*

She leaves it to her students to choose the mode of oral assessment. At the beginning of each school year, she negotiates the kind of assessment she will do with her students. The students can choose between doing oral exams on the topics they discuss in class or making presentations. In the form I observed, which was a fifth form, the students had agreed on giving a presentation at the end of the term (see below).

Like T1, T3, who teaches a fourth form, also sets only oral exams.

> Ich mache das nur auf mündlicher Basis. Ja und prinzipiell zu den Noten trägt vor allem die Mitarbeit bei, das ist ganz klar, wie sie sich in der Stunde verhalten, wie sie mittun.
>
> *//I only do it orally. Yes, and it's mainly participation that counts for the marks, that's quite clear, how they react in the lesson and participate//*

T3 usually begins the lesson with a revision of the topics discussed in the previous two or three lessons. Each student has to do such a revision twice a term. Also twice

a term, the students are assessed on a more complex topic, such as e.g. "The Roaring Twenties". They can volunteer for this exam, which "clever" students usually do.

> Sie können sich zu diesen Prüfungen auch freiwillig melden, wenn sie wollen. Was natürlich immer nur wenige machen. Das sind immer nur die sehr guten Schüler, die sind clever.
>
> *//They can also volunteer to take the exams, if they want to. Which, of course, only a few do. That's always the very good students, they aren't stupid//*

If they do not come forward of their own accord, they can be called on in any lesson. Should they not be well prepared, they do get the chance to excuse themselves once a term. Additionally, when a big topic has been covered, each student has to answer a very short question on this topic, which earns them a plus or a minus depending on the accuracy and fullness of their responses.

> Wenn wir ein größeres Kapitel fertig haben, mache ich immer Kurzfragen, nenn ich halt so. Da können sie sich ein Kärtchen ziehen mit der Frage und da müssen sie halt kurz dazu Stellung nehmen und dann stellt sich eh sofort heraus, wenn sie was können, dann sagen sie 2-3 Sätze dazu.
>
> *//If we've finished a big chapter, I always ask short questions, that's what I call it. They can draw a card with a question and then they have to say something about it and then it immediately shows if they know something, and then say 2 to 3 sentences about it//*

Such short questions would be "What was the Wallstreet Crash" or what was "Prohibition", for instance.

T2's students also have to do two revisions of previous lessons and, additionally, an exam on two independent complex topics. She offers her students the choice of date when they want to go in for the exam. She tells them the range of topics she will ask about but the students do not know on which topic they will be examined specifically.

> Diese Prüfungen, die großen Kapitelprüfungen, wo sie 2 unabhängige Fragen bekommen, da suchen sie sich selbst den Termin aus und da wissen sie, ich komme z.B. am 18. Jänner dran. Welches Kapitel wissen sie nicht, aber von – bis, es ist eingegrenzt. Das finde ich nur fair und das muss auch so sein.

*//These exams on the big topics, where they get 2 independent questions, they choose the date themselves and they know it's my turn on, for instance, January 18th. Which chapter they'll be asked about they don't know, but from a certain range, it's limited. I only find that fair and think how it ought to be//*

The more complex topics the students are asked about are, for instance, the two World Wars, the Holocaust, the Cold War, or the political and legal system in Austria.

The frequency and length of the oral exams that T4 gives in her fourth form are similar to the exams in T2's and T3's forms.

There is a striking difference, though between the questions asked by T2 and T3 and T4: T3 and T4 ask only for facts in both lower and upper secondary education. T2, however, does not solely test recall of dates and facts but frequently asks thought-provoking questions.

Example 1

T2: an additional question, but I think you can answer it. We say, mum, today that there's something like a Cold War.

S: definitely
T2: Can you support your opinion?
S: yeah, xxx shows its power with a new bomb
T2: yes, are there any other places where there's high danger?
S: Yeah, Pakistan, India, Iran
T2: Yes, very nice ….

**Presentations**

Two teachers in my study use presentations as a form of assessment. However, only one expects the students to deliver their presentations in the target language.

As mentioned above, T1's students decided on giving presentations at the end of the term. They selected topics, usually from history or archaeology magazines, and

discussed them in class. Three or four times a term, they went to the computer room, where they searched for additional information on their topics. There they worked in groups with the teacher as adviser. Successful teamwork earned them a "plus". Each student who did a presentation was rewarded with a "Sehr gut". All the students but one gave a presentation; he was assessed as "Genügend".

> T1: Es gibt klare Spielregeln für die Prüfung: Die Schüler müssen Referate halten, deren Themen sie frei wählen können. Sie suchen sich die Themen aus den Zeitschriften, arbeiten mit Buch, Lexika und Internet und erarbeiten sich die Referate selbst. Wenn sie sich an die Spielregel halten und das Referat halten, bekommen sie ein Sehr gut, wenn nicht, müssen sie sich mit einem Genügend zufrieden geben. Ich denke, das Referat ist für die Schüler eine Herausforderung.
>
> *//There're clear rules for the exam. The students have to give presentations, whose topics they can choose freely. They choose the topics from magazines, work with books, dictionaries and the Internet, and prepare the presentations themselves. If they follow the rules, they get a "Sehr gut", if not, they must make do with a "Genügend". I think that the presentation is a challenge for the students//*

In T3's form, the students do group work with material provided by the teacher once a school year. This group work is always carried out in German and focuses on German or Austrian historical topics, such as "Leben im Nationalsozialismus" *(Living under National Socialism)*. The students usually have to summarise an article given to them by the teacher and prepare a presentation. The teacher is particularly concerned that each of the students contributes to the work as in the following lesson she nominates a student who has to do the presentation of the group's summary. This presentation is done in German.

Regarding the criteria according to which presentations are assessed, the presentations given by T3's student are part of participation in class. They earn students a plus or minus, depending on their performance.

In T1's form the students' presentations at the end of the term functioned as summative assessment. There were no specific criteria laid down by T1, which should be fulfilled by the students. The sole criteria for assessing the presentations were 'done' or 'not done'. She made no distinction whether the presentations were

good or bad, neither did the students get any kind of feedback afterwards. Linguistic problems or errors were not attended to by the teacher. Interestingly, pronunciation errors created some kind of unrest and dissatisfaction among the listening students. When, for example, one student offered a variety of pronunciations of the "Phoenicians" in his presentation, one student remarked loudly:

> Könnt ma uns jetzt amol einigen auf a Aussproch? Oder könnt einer sogn wies wirklich haßt?
>
> //couldn't we agree on one pronunciation? Or could one tell us how it's pronounced correctly?

I would say if students get restive on such as issue, it might be necessary to give them a pronunciation model quickly.

**Written exams**

T2, who teaches an eighth form and T4, who teaches a fourth and an eighth form, both give either oral or written exams. The reason why they also set written tests is constraints on time. T2 usually gives written tests in the summer term because it is shorter than the winter term due to the many religious holidays. Besides, T2 thinks that both formats are beneficial for the students' further education:

> Und ich denke mir auch als Vorbereitung für die Unis oder FHs, auch dort gibt es zwei unterschiedliche Prüfungsmodalitäten: das eine ist eben das Prüfungsgespräch und auch sehr viele schriftliche Tests, und ich denke, und ich denke, beides kann man sehr gut üben.
>
> *//And I think as a preparation for university and the University of Applied Sciences, there're also two different forms of exams: the one is the interview and there're also very many written tests, and I think, and I think, both can be practised very well//*

For T4 it is the number of students that makes her opt for written exams. In her eighth form there are more than 30 students with only one lesson a week, so she cannot possibly find time slots in the lessons for testing all the students orally.

The written tests administered by T2 and T4 are similar in their design. They comprise four to five questions, which can be answered with one sentence or just a name or date, and two more complex questions on which the students have to elaborate. T2 offers the students four questions from which they have to choose two.

> T2: Der Test schaut immer so aus, dass ich 4 oder 5 Fragen stelle, die in einem Satz zu beantworten sind, oder manchmal nur 2 Namen oder 2 Verträge oder so weiter, zum Beispiel ‚name the two treaties which were most important after WWI', und da erwarte ich mir, dass St. Germain und Versaille, und da brauchen sie nur mehr die 2 Wörter einsetzen. Dann gibt es immer wahlweise 4 große Fragen, wovon sie zwei in einer Art Essay, na ja Essay wäre zu viel, beantworten müssen. Da bekommen sie 4 unterschiedliche Themen wo sie quasi zusammenfassend das eben beschreiben müssen. Es sind umfassendere Fragestellungen.
>
> *//The format of the test is that I ask 4 or 5 questions, which can be answered with one sentence, or sometimes only 2 names or treaties, and so on, for example, name the two treaties which were most important after WWI', und there I expect them to answer with St. Germain und Versaille, and they only have to fill in the 2 words. There's a choice of 4 longer questions, out of which they have to answer 2 in a kind of essay, well, calling it an essay would be going too far. There they get 4 different topics, which they have to summarize in the answer. These are more comprehensive questions//*

**T2's criteria for the assessment in written exams**

When formulating the questions for the written exams, T2 lays down the criteria that should be met by the students to achieve a "Sehr Gut". She then assesses the students' achievement in the content area according to these criteria. For the correct answer of the short questions the students get two points. Here the criterion is "answered" or" not answered".

For a full answer to the more comprehensive questions they can score between 12 and 15 points. Such a comprehensive question in a written exam might be: "The way to WWI. Describe the events that triggered the war". She expects her students not to mention solely the assassination of the heir to the throne but list and elaborate on basic problems, such as the situation in the Balkans or the role of Russia, for example.

T2: So zwischen 12 und 15 Punkten pro Aufsatz, und dann schreibe ich mir vorher schon auf, wenn ich die Fragestellung formuliert habe auf, welche Punkte ich für ein „sehr gut“ erwähnt haben möchte, schreib ich mir eben auf, z. B. the way to WWI, describe the events which triggered the war, schreib ich mir eben auf, net nur die assassination tatsächlich, sondern z. B. basic problems in the Balkans, the role of Russia, und das schreibe ich mir stichwortartig auf. Diese 4 oder 5 Punkte sollte jemand erwähnen, wenn ich finde, dass das die gesamte Punkteanzahl ist.

*// Between 12 to 15 points per essay, and then I write down in advance, once I've formulated the question, which points I want them to have mentioned to earn a "Sehr gut", I write down, for instance, the way to WWI, describe the events which triggered the war, not only the assassination, but for instance, basic problems in the Balkans, the role of Russia, I list key points. These 4 to 5 points should be mentioned in order to score full marks//*

For each point that is not mentioned she deducts points. Yet she wants her students to demonstrate what they have understood rather than to memorise facts. With respect to the question about WWI, she wants her students to understand the political circumstances that gave rise to the war.

T2: Und dann gibt's halt Abzüge, aber da sind nicht jedes kleine Detail das drinnensteckt, sondern ich versuche das sehr global, was erwart ich mir, dass jemand bei dieser Fragestellung wirklich sagt, dass sie bei den Ursachen, die zum 1. Weltkrieg geführt haben, nicht nur die Ermordung des Thronfolgers, sondern dass dieser Imperialismus, der vorher schon war, sondern das Muskelspiel für die Vorherrschaft in Europa. Und dann gibt's halt Abzüge. Wenn jemand nur wischi-waschi schreibt, dann ziehe ich ab.

*//And then there're deductions, but not each minute detail, but I try to do it very globally, what do I expect somebody to be able to say in an answer to this question, that they say something about the causes for WW I, not only the assassination of the heir to the throne but that that imperialism, which had already been there, the muscle play for the hegemony in Europe. And then there're deductions. If somebody only writes bla - bla, then I deduct//*

### 6.3 Choice of code in assessment

In CLIL, the language of instruction is English and the majority of topics are taught and discussed in English. Austrian civics, however, is usually presented and dealt

with in German; historical events from German and Austrian history are either taught in German or English.

> T2: Politische Bildung schon auf Deutsch. Gerade wenn es Österreich betrifft, halte ich es für sinnvoll, dass man es zunächst einmal auf Deutsch erklären kann, was ist der Nationalrat und der Bundesrat. Österreichische und deutsche Geschichte mache ich schon auf English.
>
> *//Civics I do in German. Especially, when it's to do with Austria, it makes sense to be able to explain first in German what the Lower and Upper Chambers are. Austrian and German history I do in English//*

> T3: der Großteil ist in Englisch als Arbeitssprache, zu zwei Drittel auf alle Fälle, aber gewisse Kapitel, auch österreichische Geschichte mache ich auf Deutsch.
>
> *//The majority is done in English as the working language, for about two thirds in any case, but certain chapters, also the Austrian history, I do in German//*

T4 teaches Austrian and German history in English due to the suggestion of the school inspectorate.

> Ja, das ist ein Vorschlag von der Frau Landesschulinspektor. Sie will es so. Es ist okay für mich. Mache ich auf Englisch. In der deutschen Geschichte ist relativ sehr viel gerade im 20. Jahrhundert mit internationaler Geschichte verknüpft, dazu gibt es ohnehin sehr viel auf Englisch und österreichische Geschichte, bzw. Politik usw. das muss man sich heraussuchen und umschreiben.
>
> *Yes, that's a suggestion of the school inspector. That's how she'd like it. I do it in English. In the 20th century German history, there're quite a lot of links to international history, there's a lot of material in English, and the material for Austrian history and politics I've to find myself and translate//*

According to the legal regulation (see chapter 2.2) teachers are not allowed to insist on answers in English in exams. Consequently, students are free to choose between their mother tongue and the target language. Interestingly, this issue is never a point of discussion. In their exams the students follow the language of instruction: the topics that have been taught in English are answered in English while Austrian

Civics and German and Austrian history, if taught in German, are responded to in German.

> T4: so wie sie es serviert bekommen, möchte ich vorsichtig sagen, so reden sie dann auch bei der Prüfung. Und die Unterlagen haben sie auch auf Englisch. Normalerweise, wie gesagt, in der 4. überhaupt, was ich Deutsch unterrichte dort machen sie es zum Teil auf Deutsch und wo es ihnen auf Englisch serviert wird auf Englisch und die Oberstufe sowieso nur auf Englisch. Da ist überhaupt keine Diskussion darüber.
>
> *//The language they use in exams is the language it was administered in, to put it carefully. And they have the material in English. Normally, as I said, mainly in the 4th form, they partly do in German what I teach in German, what they get in English is done in English, and in the upper forms only in English. There's no discussion about it.*

> T3: [...] österreichische Geschichte mache ich auf Deutsch. Und das ist dann in der Prüfung auch auf Deutsch, natürlich, sie reden immer so wie sies hören.
>
> *//I do the Austrian history always in German And that's the language they use in the exam, too. Obviously they speak in the language they hear it in//*

T4 said that only one student in her career as a CLIL teacher preferred to do the exam in German. He was on the verge of failing the class and felt safer to do the exam in German.

> T4: Eventuell wenn es darauf ankommt, dass es sozusagen, vorsichtig gesagt, wenn es einen Entscheidungscharakter hat, dann kann es sein – wenn zwischen 4 und 5 – dann kann es schon sein, dass sie sagen ich würde mich auf Deutsch leichter tun – das hatte ich bisher nur einmal.
>
> *// Ultimately, when it comes to the crunch, to put it carefully, if it's an important decision, if it's between 4 and 5, it could be that they say they'd rather do it in German. I only had that once so far//*

In T2's eighth form one student opted for German in the exam because he was going to do history in German in his school-leaving exam.

Example 2:

T2: The start of the Cold War. Describe the situation in Europe after WWII

S: Frau Professor, darf ichs auf Deutsch machen, weil ich mach die Matura auch auf Deutsch.
//*May I do it in German because I do the final exam also in German//*
T2: Ja, wenn das funktioniert, dass du das was auf Englisch
//*Yes, if it works that you what is in English//*
S: Ja, ich hab mirs da auf Deutsch zusammengefasst
//*Yes, I've summarized it in German//*
Naja, also der kalte Krieg herrschte zwischen Sowjetunion und Amerika...
*//Well, the Cold War was between the Soviet Union and America....//*

One student in T1's 5th form switched from English to German due to lack of language proficiency. She showed an intensive interest in the subject she discussed in her presentation and employed German for two different reasons: firstly, by translating specific terminology into German, she made sure that all her classmates could understand what she wanted to put across. Secondly, she switched over to German because her competence in English was not sufficient enough to present all her findings. She had gathered a considerable knowledge of her subject but was unable to express it.

S: Frau Professor, das muss ich jetzt auf Deutsch sagen. Ich muss noch soviel dazu sagen. Das hab ich in einem Artikel gelesen, aber ich weiß nicht, wie man das auf Englisch sagt.

*//Frau Professor, this I must say now in German. I've got so much to say to it. This I've read it in an article but I don't know how to say it in English//*

**Code-switching**

If students have difficulties with vocabulary in the exams, they usually hesitate and give the expression in the mother tongue. They do not need to ask what the corresponding word in English is. For the teacher their hesitation and the word in German are signals enough to help. In most of the cases, it is only one word as in the following two examples.

Example 3:

S: […] and the second one, the hydrogen bomb, there is one atomic bomb to … zünden?…
T: to trigger
S: to trigger the hydrogen, and the hydrogen atoms are pressed together…

Example 4:

> S: and so … überbrücken ?… the distance between
> T: bridge the distance
> S: bridge the distance between America and Russia without….

In either case the student who does not know the particular word hesitates, names it in the mother tongue in a questioning tone, pauses, and looks at the teacher for help. The teacher supplies the word and the student then takes it up himself and continues in English.

In the example below the student is trying to re-formulate the expression "loss of identity", which she had in her mind but was not available at that very moment, to clarify her answer and put it across in the exam.

Example 5:

> T: in the aftermath of WW I and WW II, what was the main difference, what would you say?
> S: dass gwusst haben wer sie sind nach dem 2 Weltkrieg, dass nicht so schlimm war. Nach dem 1. Weltkrieg habens nicht gwusst wer sie sind.//that they knew who they were after WWII, that it wasn't so bad. After WWI they didn't know who they were//
> T: why didn't they know it?

The wording of her re-formulation, however, proves that she had perfectly understood the meaning of the terminological expression.

According to the teachers, code-switching is fully accepted and does not influence the grade, which is clearly expressed by T3.

> T3: Es kommt hie und da ein deutsches Wort einmal vor und da hilft man ihnen weiter und ich bezieh es überhaupt nicht mit ein. Es muss schon entweder englisch oder deutsch sein, aber es sind ja nur einzelne Worte, aber wenn sie wissen was das auf deutsch heißt, dann ist es mir auch recht oder es hilft irgend jemand anderer weiter, der aufzeigt und sagt was das heißt und es geht dann schon weiter auf englisch. Das Sprachliche beziehe ich überhaupt nicht ein in die Prüfung, solange es verstanden wird, inhaltlich klar ist. Ich meine wenn das so schlecht wäre, dass da nichts mehr verstanden würde, aber

das ist ja meistens nicht der Fall. Das sprachliche zählt überhaupt nicht, die können Fehler machen soviel sie wollen. Für die Note gilt nur der Inhalt.

*//Every now and then there's a German word, and then you help them, and I don't take it into consideration. It should really be either English or German, but it's only single words, but if they know what it is in German, then it's o.k. for me, or else somebody else helps by putting up their hand and saying what it is, and off we go in English. I don't take the linguistic element into consideration at all in the exam as long as it's understood and the content is clear. I think if it was so bad that you couldn't understand anything, but that's generally not the case. The language component doesn't count, they can make mistakes as many as they want. Only the content is important for me//*

### 6.4 Error correction

As an English teacher in adult education I have experienced a difference in learners' language behaviour in recent years: in my courses, the students having undergone CLIL instruction at school speak more confidently and fluently than those having done solely the traditional English classes. The former CLIL students claim that their confidence in and fluency of their speech is a consequence of not having been corrected in their linguistic performance and not examined on their language proficiency. They could speak freely and without strain about the content of the subject and thus gained confidence in speaking in the foreign language.

Dalton-Puffer states that students' "self-confident and self-evident use of the foreign language and its ultimate appropriation by many CLIL learners" derives from "CLIL classroom discourse" that "provides a space for language learners where they can act in a context that is not geared specifically and exclusively to foreign language learning (i.e. it has a 'real' purpose) but at the same time is predefined and pre-structured in significant ways by being educational-didactic and taking place within the L1 matrix culture" (2005:248). I go along with this argument but I think that students are more highly motivated by the conviction that their errors are not assessed.

**Linguistic errors**

The teachers in this study hold different opinions about the correction of linguistic errors. On the one hand, it seems that they feel rather liberated from the burden of correcting language errors when teaching CLIL classes. As long as the errors do not impede or impair understanding, the teachers claim, they ignore them.

> T1: Nein, weil ich denk, wir sind sowieso daraufhin trainiert, dass wir nur Fehler hören und Fehler sehen. Ich mag sie sprechen lassen und wenn es inhaltlich passt und wenn ich den Inhalt versteh, ist die Sprache nebensächlich.
>
> *//No, because I think we're only trained to see and hear mistakes. I want to let them speak and if the content is correct and I can understand it, the language is irrelevant//*

On the other hand, the teachers think that it is their duty to correct the errors. T3 informs her students at the beginning of CLIL that she will correct errors but will not assess them as part of the exam.

> T4: Wenn es extreme Fehler sind, ignorieren kann ich es nicht, im Raum stehen lassen kann ich es auch nicht wirklich, das kann ich als Lehrerin ja doch nicht. Ich bessere es aus und das wars. Wird nicht beurteilt, wird auch nicht weiter besprochen, das ist dann Sache des Englischunterrichtes. Da trenne ich sehr.
>
> *//If there're serious mistakes, I can't ignore them, I can't just let them just happen, I really can't do that as a teacher. I correct it and that's that. Is not assessed, will not be discussed, that's something to be handled in English lessons. I make a sharp distinction here//*

In the following example a student of the 8th form was asked about the political situation in Austria after WWII.

Example 6:

> […] to accept Karl Renner's government and, ahm, but finally they … authorotised
> T: authorised
> S: they authorised his government and then, ja, I think they wanted to give Austria a chance to recover.

The student's lexical error is corrected by the teacher; repeating the correct word, the student continues her speech.

The following two examples are also taken from an oral exam of a student in the eighth form. The student was asked about the climax of the Cuba crisis, in particular about the role of the United States and Russia in this conflict.

Example 7:

> T2: […] what do you think of the role of the US before Fidel Castro? […] Were they concerned about Cuba?
> S: it was unimportant for them, they were only **interested for** their own countries.
> T2: mhm, can we say they exploited Cuba?

Example 8:

> T2: From what we've heard so far, would you think that one of them would have pushed the button?
> S: I think both were frightened because if, if… **they … would did so**
> T2: if **they had done so**
> S: the other **one also have done** so
> T2: and so you think…

In example 7 the student's linguistic error in form of a wrong preposition is passed over by the teacher as it does not impair understanding.

In example 8 the student is asked a hypothetical question; he should express his opinion on the potential use of nuclear missiles in the Cuba crisis. Although the teacher provides the grammatical structure of the conditional in her question, the student is unable to make use of it in his answer. The teacher immediately provides the correct structure to help the student out of his linguistic "predicament". However, the student obviously does not feel obliged to repeat the correct structure; rather he seems to feel relieved to have overcome this bad situation. He continues answering, again with a wrong structure. This time then the teacher deliberately ignores the mistake and turns to the next question.

**Factual errors**

Factual errors were corrected on the spot by the teacher. Interestingly, these very rarely occurred in the exams I observed. The below-mentioned example is indeed the only one I detected in all the recordings. Normally, the students did not say anything if they were not sure about the answer as illustrated in example 10.

Example 9:

T: and who were the customers in the gambling casinos?
S: ah, ….the Cubans
T: no, no! American business men

Example 10:

S: ich war erst dran, gibts jemand, der noch nicht dran war
*(I've just been examined. Isn't* there *anybody who hasn't been examined yet?)*
T: na, anything you can remember
S: …
T: what do you remember
S: …

One suggestion why factual errors hardly ever happen might be that the line of thought is shaped by the teacher, taken up and parroted by the students. The student's wrong answer in example 9 might support this assumption. Who gambled in Cuban casinos is perhaps not an important issue in the Cuba Crisis and thus might not have been discussed in the previous lessons. It has to do with logic and might reveal that the student had either not followed the discussion or not properly grasped the underlying implication about US-Cuban relations at the time of the Cuba Crisis.

## 6.5 The role of language

The following observation on the role of language in assessment is mainly based on the oral exams embedded in the lessons of T2 and the presentations of T1's students. While T1 reserved the last three lessons of the term for her students' presentations, which served as summative assessment, T2 carried out assessment throughout the term. For this reason, I had – and was allowed – to observe the whole lessons, as in

her form participation in class, revision of previous lessons and an exam on specific topics summed up to the final grade the students received.

It should be pointed out that this investigation was conducted in an eighth form. The students are very near to their "Matura" (school-leaving exam), where they, ideally, will have reached level B2 of CEF[16]. At this level learners are supposed to

> […] understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

In other words, the students are expected to understand the topics they are presented with in this form. However, the students do not have just one area of specialisation. But there should be no special need for the teacher to work on the core language or to modify her output by, for instance, slowing down or repeating or reformulating information.

With regard to the students observed in this study I could form an opinion about T2's students' language proficiency in the 8th form. However, I was allowed access only to the last three lessons of the term in T1's 5th form, in which the students were assessed on the presentations they gave (see chapter 6). As these presentations were on the whole scripted, and were partly read, I could not identify the students' "own language". T1 did not help to clarify the matter: neither did she correct her students nor did she give them immediate feedback on their attainments (see below).

[16] http://www.ceftrain.net/ p. 24

## 6.6 The role of language in assessment

The most relevant aspect of assessment in CLIL in my investigation is the role that the target language plays in students' performance. While it has been explained that the effect of error correction per se is probably over-estimated (see chapter 6.4), the impact of whether errors determine assessment or not needs to be considered separately. It is an important aspect of answering the question which role the language plays in assessment.

In the Austrian CLIL classroom it is the content that is assessed while language "must be left out" in assessment (see below). Interestingly, the question as to where this "demand" comes from could not be answered satisfactorily by any of the teachers; it is a matter of course. One teacher (T2) said that she might have heard it in one of the inofficial meetings about CLIL she had attended (see chapter 6.1, cf. chapter 3.2).

For the students in my study there is no doubt that they are only tested on their content knowledge, which is expressed in the following two statements taken from communications with students:

> S: das wissen wir, dass wir nur über den Inhalt geprüft werden.
>
> *//we know we're only on assessed on content //*
>
> S: das sagen uns die Lehrer gleich am Anfang, dass die Sprache keine Rolle spielt. Den Inhalt müssen wir können, wie wir den rüber bringen ist egal.
>
> *// we're told by the teachers right at the beginning that the language doesn't play a role. We must know the content. How we transmit it, doesn't matter//*

### The interviews with the teachers

As far as the teachers' statements are concerned there is a glaring discrepancy between what is required in the literature (see introduction) and what is consistently maintained in the interviews that I carried out with the teachers in my study. The teachers make no demands for new methods or tools, nor do they express a feeling of uncertainty about the role of language or indicate unwillingness to talk about it (see

chapter 1.1). They state clearly and with conviction: **language is ignored in assessment**.

T1: nein Sprache bewerte ich überhaupt nicht, mir kommt es nur darauf an, was sie mir erzählen, ob halt der Inhalt richtig ist.

*//no, I don't assess language in any case, I'm only concerned with what they are telling me, if the content is correct//*

T2: nur der Inhalt, nein, die Sprache dürfen wir, glaub ich, gar nicht beurteilen.

*//only content, no, the language, I think, we aren't even allowed to assess//*

T3: Mich interessiert nur der Inhalt. Sprache steht außen vor, weil es eine Geschichtestunde und keine Englischstunde ist. In diesem Fall prüfe ich ihr Geschichtewissen und nicht ihr Englisch Können.

*// I'm only interested in the content. Language is ignored because it's a history lesson and not an English lesson. In this case I test their knowledge in history and not their competence in English//*

T4: Das Sprachliche beziehe ich überhaupt nicht mit ein in die Prüfung..Das Sprachliche zählt überhaupt nicht, die können Fehler machen soviel sie wollen oder deutsche Wörter verwenden. Für die Note gilt nur der Inhalt.

*//I don't include the language element in the exam. The language element doesn't count. They can make mistakes as many as they want or use German words. For the grade it's only the content which counts//*

**Assessment of oral exams**

The teachers' absolute belief and assertion that they pay no attention to the language in assessment, however, was considerably shaken, indeed turned upside down, at my very first encounter with assessment in CLIL. I was invited by T2 to come to a CLIL class in order to become familiar with the practice and procedure followed in class. As I only wanted to introduce my project and gather information about how to organise my observation, I was neither equipped with the camcorder nor the voice recorder. Therefore, this informal, but for the development of my study crucial, encounter can only be rendered as an anecdote, which then, however, was reflected upon and discussed in an extensive interview with T2.

The topic of the lesson I took part in was the Holocaust and two students were nominated to revise briefly the history of the Jews, which had been dealt with in the previous lessons. Content-wise both students answered the questions correctly and satisfactorily for the teacher; each of them covered all the relevant facts (confirmed by the teacher). Linguistically, though, there was a major difference in their performance. S1 spoke quickly and fluently, using a number of linking words. She expressed absolute self-confidence in her behaviour; sitting upright, looking straight into the teacher's eyes, she seemed to enjoy the situation. Her language performance, according to my impression and the teacher's confirmation, corresponded to the level expected by a student of the 8th form (B2 of the CEF). S2, in contrast, spoke less quickly and fluently, worded his thoughts in half-sentences, paused several times, and listed the facts rather than putting them in coherent speech. His posture assumed less confidence in his knowledge as well as less command of language than the student preceding him.

When asked about the grades she had given to the students, the teacher told me that, since it was an informal exam, she was not allowed to give grades but symbols. Translated into grades these would mean a 1 for S1 and a 3 for S2. She immediately justified her grading as follows: "well, you see, **her language is so much better"**. To the teacher's utter surprise, the language element, whose role in assessment had been so firmly denied previously, was there. Moreover, it finally decided the grade the two students achieved.

**The interview with T2**

In the interview, which followed some time later, the teacher commented on the role of the language in the exams and stated first that she welcomed the research project, in particular for her own development (which might encourage teachers to open their classroom doors for further research):

> Zuerst nur Inhalt, Sprache nein. Das war das Interessanteste vom ganzen Projekt. Es ist generell sehr interessant. Ich würde es öfters gerne haben. Einfach ein Feedback von Leuten von außen, das ist nicht ganz einfach aber sehr sinnvoll und weiterführend. Auch ein eigener Zensor und hilft in der Weiterentwicklung des Lehrers.

*// At first only content, language no. This was the most interesting part of the whole project. It is generally very interesting. I'd like to have it more often. Just feedback from people from outside, which is not easy but something that makes sense and helps you develop, a personal "censor", all that helps//*

Previously, the teacher had been absolutely convinced that she only assessed language, which she had also passed on to her students.

> Und habe mir gedacht gut, ich habe das auch immer den Kindern so erklärt, es wird die Sprache nicht bewertet und da braucht ihr keine Angst haben, es wird nur der Inhalt und habe mir gedacht, das ist so und mache ich auch so, war da wirklich sehr überzeugt.
>
> *//and I thought, and I also explained it to my students, that language is not assessed, and they needn't be afraid, it is only content. and I thought that's the way it is and I do it this way, and I was really very much convinced of this//*

The completely unexpected outcome made the teacher reconsider and reflect on her own assessment practices:

> Und wie wir dann begonnen haben, oder wie sie nachgefragt haben, so quasi was war das jetzt, was hast du dem gegeben, welche Noten dem F. (S2) und G. (S1), die verbal sehr gut ist, ein flüssiges Englisch spricht und F.(S2) war der, der ist ein blitzgescheiter Kerl, nur ist er halt in der Sprache nicht so da. Sie hat ein eingeringeltes + (Plus) bekommen und er hat, sag ich einmal nur ein + bekommen. Wie sie dann gefragt haben und der Inhalt? Na ja, das stimmt schon, aber inhaltlich hat der eigentlich genauso viele Dinge auf den Punkt gebracht und erwähnt, und da habe ich zum Nachdenken begonnen und, na ja, und bin mit mehr Fokus hineingegangen und habe mich selber ein bisschen mehr beobachtet und hinterfragt, beim Beurteilen auch so, was habe ich jetzt beurteilt?
>
> *//and when we began, or when you asked me, so quasi what was that, which grades did you give F. (S2) and G. (S1, who is verbally very good and speaks fluent English, and F. (S2), who is very intelligent, but his language isn't so good. She got a 'plus in a circle' and he, let's say, only a 'plus'. When you asked then and the content? Well, that's correct, content-wise he has come up with just as much. Then I began to think about it and, well, focussed more on it, watched myself more and questioned myself while assessing about what it was I had assessed//*

The teacher pointed out that she had always thought that she only assessed content and she never prefers some students to others:

> Ich war auch felsenfest überzeugt, ich schaue nur auf den Inhalt und ich bin wirklich niemand der wen austrickst oder so mit Sympathien, das ist wirklich nicht so mit Sympathien, dass ich wen bevorzuge.
>
> *//I was firmly convinced that I only look at the content and I'm really not a person who is unfair to some or favours others, that's really nothing with sympathies, that I prefer someone//.*

Only through the research project had she become aware that the overall impression of the student's performance, comprising language proficiency, pronunciation and the student's body language, determined the grade they could obtain:

> Man lässt sich von der Sprache verleiten. Weil da klingt alles rund und toll und der macht keine Pausen und dann ist die Aussprache noch gut und dann klingt es auch schon einmal viel besser, dann hat er auch eine gute Körperhaltung und schaut einem in die Augen, ist ruhig, gelassen und dann glaube ich ist es schon ein verfälschtes Bild.
>
> Das ist mir wirklich erst bei diesem Projekt quasi bewusst geworden. Und ich hab mich selber ausgetrickst oder austricksen habe lassen. Was man aber eigentlich wird, weil wer geschickter mit der Sprache umgehen kann und das dadurch auch anders wirkt, bekommt eine bessere Note und weil ein anderer, der holpriger oder viel holpriger spricht, der hätte vielleicht 5 Minuten länger gebraucht und hätte es auch hinbekommen, aber es klingt nicht nach 1 und ich hätte ihm wahrscheinlich auch keiner 1 gegeben. Das war auch die große Lehre für mich, dass es so leicht dahingesagt ist, die Sprache wird absolut nicht benotet und man tut es doch, das ist das große Aha-Erlebnis in allen Fächern. Ich bin auch tief davon überzeugt, das passiert auch in der Muttersprache Deutsch.
>
> *// one is misled by the language. Because everything sounds smooth and great and he doesn't pause and then the pronunciation is very good, and then it sounds so much better, and then he holds himself well and looks into your eyes, is calm and relaxed, and then I think the impression is distorted.*
>
> *I only became aware of this during this project. And I have outsmarted myself or have been outsmarted. What actually happens, because someone who is better at the language and thus makes a different impact, is that he gets a better mark and the other one who speaks awkwardly, he'd have needed five more minutes and he'd have managed but it doesn't sound like a 1, and I probably wouldn't have given him a 1. That taught me a lesson that it's easily said that language is not assessed and yet you do it. That was the moment of insight for me in all subjects. I'm deeply convinced that this also happens in the mother tongue//*

The teacher's remark that "das passiert genauso in der Muttersprache Deutsch". *("the same happens in the mother tongue German"*) is perfectly correct. It does happen in the mother tongue German, too. In fact, it happens in all languages. And yet, there is a major difference. If the same exam situation as described above were transferred into a history lesson in the mother tongue, the results, due to the factors explained above (see chapter 2.5), might have been the same. However, it is most unlikely that they would have been justified by saying one student was more eloquent than the other. In the mother tongue, the language remains "invisible". In CLIL, however, it becomes "visible" (see chapter 4.1).

Moreover, the target language per se adds value to the student's performance, as was expressed by the teacher when talking about her students' participation in class (see chapter 6.2).

> T2: [...] noch dazu, wenn es in Englisch gesagt wird.
>
> *//[...] especially, if it comes out in English//*

When explaining the different components of assessment, T2 emphasised that students' contribution to topics being discussed is given more weight if it is spoken in English. Interestingly, the teacher gave this statement after she had talked extensively about the role of the language in the assessments she had carried out in CLIL. The value of the target language seems to be firmly anchored in the teacher's mind, on the one hand; on the other, she claims to leave it out of account.

**T2's comments on transcribed oral exams**

In order to clarify the teacher's system of grading further, I asked the teacher to comment on the 9 oral exams that I had recorded and transcribed. Focusing on the content, she emphasised that students who were able to answer questions that call for explanation or opinion are more likely to get better grades than those who recall and reproduce mere facts. In the oral exams that she administers she not only asks content questions that have been dealt with in the previous lessons, but adds question

such as “can you explain?”, “what does it mean ?”, “what do you personally think?” (see chapter 6.2).

Example 11:

S: both had nuclear weapons but the USA had more of them
There were two kinds of atomic weapons, the hydrogen bomb and the atomic bomb
T: mhm, can you **explain** how they work?
S: the .. atomic bomb has two amounts of uranium, they get together .. and till they have the critical mass and then explode, and create huge energy when they explode
T: mhm
S: and the second one, the hydrogen bomb, there is one atomic bomb to .. zünden?,
T: to trigger
S: the hydrogen, and the hydrogen atoms are pressed together and create also (xxx)
T: mhm, which bomb is stronger, atomic bomb or hydrogen?
S: hydrogen.
T: good

Example 12:

T: **what does** neutral **mean**? We haven’t talked about it, so this is a question to everybody. **What does it mean** that Austria is a neutral state?
S: we’re not allowed to enter in a war or help anybody
T: and also not into the Warsaw Pact and NATO, which still exists
S: and, and at the 25th of October the last troops left Austria, and the next day, the 26th of October the declaration of neutrality was signed, and this became our national holiday, of Austria.

Example 13:

T: an additional question, but I think you can answer it. **What do you personally think?** We say, mhm, today that there’s something like a new Cold War
S: definitely
T: **Can you underline your opinion?**
S: ja, xxx shows its power with a new bomb
T: yes, are there other places where there’s high danger?
S: Yeah, Pakistan, India and Iran, USA
T: Yes, very nice J., das macht das Sehr gut aus, so zu denken.

The oral exam, of which example 13 is the concluding part, and the corresponding teacher's comment (see below) is particularly interesting. The student's response to the question for his opinion and his subsequent considerations was praised by the teacher and earned him a "Sehr gut". He fully satisfied her expectations when he "underlined his opinion" and presented facts that had not been discussed in the lessons but collected extramurally. That means that the impression she gained from the student at the end of the exam, when he showed knowledge beyond the topical competence required, decided the grade and obviously made her overlook that the student had not performed too satisfactorily throughout the exam.

A closer look at the whole exam underpins this assumption. The student was asked about the political situation in Europe after WW II.

Example 14:

T: J.
S: the conference, es
T: sorry?
S: the conferences?
T: no, we'll come to that, okay,… start with the conferences
S: there were two conferences, … the conference of Yalta and of Potsdam, and, ja, in Yalta they made three main decisions
T: mhm
S: first ,ahm, ahm, ..., ...,...
T: ((laughs)) mhm, don't get nervous, …okay, Yalta conference, Potsdam conference, what were the results? Once again
S: first the borders of Poland were created
T: mhm, not created, renewed, I'd say
S: the xxx, in Yalta they founded the UNO
T: yes, did they invent the United Nations anew?
S, no, no, first it was the xxx Nations
T: ja, very nice, okay, good, other results?
S: the zones of occupation were founded, … four zones,
T: okay, which countries were taking part
S: America, France, Great Britain and Russia
T: mhm, okay, good. Then after a long time of tensions, of showing muscles to each other, the two nations founded two new pacts, which pacts?
S: …
T: well, throughout the years, there were always communist countries against capitalist countries, and their heads
S: Warsaw Pact
T: yes, Warsaw Pact countries are which countries?
S: east
T: yes, eastern countries, and the NATO are
S: the NATO are 12 countries
T: do these pacts still exist?

S: ja,
T: would you say, they're important?
S: ja, the NATO is very important

[…]
T: an additional question, but I think you can answer it. **What do you personally think?** We say, mhm, today that there's something like a new Cold War
S: definitely
T: Can you underline your opinion?
S: ja, xxx shows its power with a new bomb
T: yes, are there other places where there's high danger?
S: Yeah, Pakistan, India and Iran, USA
T: Yes, very nice, J., das macht das Sehr gut aus, so zu denken.

The student received the grade "Sehr gut" for this exam. However, the performance he delivered was not altogether convincing. His speech was fairly slow and occasionally hesitant, especially at the beginning. The teacher had to encourage him and guide him through the exam with prompts such as requesting further information or offering extra information when he did not answer at all. The student's answers were very short and linguistically rather simple, his range of vocabulary being fairly small. He enumerated single facts, at times he only provided a list of names, e.g. America, France, Great Britain and Russia, at others he gave merely an affirmative answer to the teacher's question such as 'ja', or 'ja, the NATO is very important'. Throughout the entire exam the teacher did more talking than the student.

After reading the transcript, the teacher commented on the exam as follows:

> Note: Sehr gut. Obwohl, jetzt in der Distanz, nach Durchlesen des Transkripts, hätte ich kein Sehr gut gegeben, aber Benotung mündlicher Prüfungen erfolgt nun mal aus der Situation im Hier und Jetzt. Begründung: zu Beginn eher holprig, Lehrerin muss viel Hilfestellung geben, Frage/Antwort Spiel, Schüler erzählt zu Beginn nicht selbständig – ist eher ein Kriterium für das Sehr gut – aber: ist das nun nicht auch wieder Bewertung der Sprache? (gilt auch für Prüfungen in Deutsch). Schüler kann aber alle Fragen, die selbständiges Denken und persönliche Meinung erfordern, sehr gut und klar beantworten.
>
> *//Grade; Sehr gut. Although, now some time has elapsed, after reading the transcript, I wouldn't have given a Sehr gut, but the grading of oral exams does result from the situation of the here and now. Reasons: at the beginning rather awkward, teacher must give a lot of assistance, question-and-answer-game, student doesn't speak independently at the beginning – a criterion for the Sehr gut – but: wouldn't this be again assessment of language? (the same*

*would apply to exams in German). But student can answer all question, that require independent thinking and personal opinion very well and clearly//*

Here, the teacher addresses three significant factors affecting the judgement of oral exams: oral exams are "fleeting", that is the teacher has to judge the performance on the spot and has no possibility to reconsider and change her grade. Furthermore, the student's speech lacks fluency ("eher holprig"), and, additionally, the proportion of his speech in the exam is rather low ("Frage/Antwort Spiel"). These factors, among others, are likely to skew the grades in oral exams (see chapter 2.5).

The second factor, namely the lack of fluency in speech, was also the reason why the student who did the exam in German (see chapter 6.3.) had to make do with a lower grade although he answered all the content questions to the teacher's satisfaction.

> Ich bin mir hier beim Durchlesen sicher, dass die Tatsache dass der Schüler die Prüfung in Deutsch gemacht hat, unbewusst eine Rolle gespielt hat. Es wirkt nun mal holprig, wenn ein Schüler den Inhalt, den er im Englischen gehört hat, auf deutsch wiedergibt. Damit ist es natürlich eine Beurteilung der Sprache – „mea culpa".
>
> //*When reading the transcript, I'm sure that the fact that the student did the exam in German unwittingly played a role. It does sound awkward if a student reproduces the content he's heard in English in German. This is of course assessment of language-mea culpa//*

The highest praise was earned by the student who was totally confident with regard to content knowledge and linguistic performance. Without being prompted by the teacher she spoke fluently and quickly on her own, giving correct and complex answers. Her share of speech was larger and the range of her vocabulary greater than in her peers' performances. Furthermore, she was able to answer additional questions to do with logic or asking for explanation. Her performance was met with full approval from the teacher and was awarded a "Sehr gut".

Example 15:

> T: Describe Austria after WWII
> S: Well, Austria was divided into 4 zones, it was occupied by the Soviets zone, or it was divided into the Soviet zone, the Great Britain zone, the French zone and the US zone.
> At May the 8th 1945 the 2nd Austrian republic was … established

T: Mhm
S: and so Austria became independent and so Karl Renner was appointed as the first chancellor in the 2nd Austrian republic and he then became the first Austrian president, and each zone was ruled by a command, by a commander, and they also held meetings, where mhm they discussed the situation or what they are going to do now and at first they didn't want to ahm, ahm anerkennen
T: to accept
S: to accept Karl Renner's government and, ahm, but finally they authorotised
T: authorised
S: they authorised his government and then, ja, I think they wanted to give Austria a chance to recover
T: exactly, here's a point where I'd like to interrupt and ask you an additional question. You said they wanted to give Austria a chance to recover. What **makes the difference between Austria after WWI and WWII**. I mean, there was a lot of destruction, people were suffering, what was the main difference, what would you say?
S: dass gwusst haben wer sie sind, dass nicht so schlimm war, nach dem 1. Weltkrieg habens nicht gwusst wer sie sind
[...]
S: Austria was a free, independent, democratic
T: neutral
S: neutral state and, and
T: **what does neutral mean?** We haven't talked about it, so this is a question to everybody. What does it mean that Austria is a neutral state?
S: we're not allowed to enter in a war or help anybody
T: and also not into the Warsaw Pact and NATO, which still exists.
S: and, and at the 25th of October the last troops left Austria, and the next day, the 26th of October the declaration of neutrality was signed, and this became our national holiday, of Austria.
T: okay, has become, because it's still celebrated
Excellent, really, sehr gut

The teacher's comment confirms her great satisfaction with this exam:

> Wirklich ansprechende Prüfung, sowohl inhaltlich gut eigenständig dargestellt, kann auf Interventionsfragen (zur Übung der Matura, da ist ja auch ein Dialog und kein Monolog) sehr gut reagieren, wirklich fein.
>
> *//a real pleasing exam, presents the content on her own, can respond to additional questions very well (as an exercise for the school leaving exam, there's also a dialogue and not a monologue), really fine//*

The teacher is extremely concerned about being fair to her students and keeping to her promise not to take the language into consideration in assessment, which she gave her students at the beginning of the CLIL instruction. She informed her students

about the surprising outcome of my very first observation, when she explained to me that one student received a better grade than the other due to higher linguistic skills (see chapter 6.5).

> Ich habe es ihnen aber gesagt. Ich bin so ehrlich und dann passen sie vielleicht auch besser auf, denk ich mir. Ich will auch die Schüler zum selbständigen Denken erziehen und Meinung sagen. [...] hab ich ihnen gesagt, das hat wirklich was gebracht, dass die Frau Hönig gekommen ist, und habe das von der G. und dem F. erzählt. Und wollte damit auch erreichen, dass sie mich auch ein bisschen mehr kontrollieren. [...] **und mir das wird wirklich nochmals passiert,** und die Schüler finden, im Grunde genommen war der auch gut, warum kriegt der eine schlechtere Note? [...]. Weil ich denke mir, das ist ihr Recht zu erfragen wie eine Note zustande kommt. Und ich bin wirklich gerne bereit zu begründen, warum er die oder die Note kriegt. [...] Zum Beispiel, unsere Wahrnehmung ist, dass der F. auch das Alles gesagt hat, halt im schlechteren Englisch. Und sie sagen immer die Sprache zählt nicht.
>
> *//But I've told it to them. I'm that honest and perhaps they'll become more attentive, I think. I also want to train the students to become independents thinkers and express their opinions. […] I told them that it was very useful that Mrs. Hönig had come and I told them that about G. and F. And also wanted them to keep an eye on me. […]* ***should it really happen to me again*** *and the students think that basically a fellow student was also good, why did he get a lower grade? […] Because I think they have the right to know how a grade was arrived at. And I'm really willing to give the reasons why the student gets this or that grade. […] For instance, our perception is that F. also said everything, but in worse English. And you keep telling us that language doesn't count.*

It did happen to the teacher again, as she explained in her comments on the oral exams, as for instance in her comment on the exam in example 13, and on the exam that the student carried out in German (see above). It should be pointed out that all the exams were administered and assessed by the teacher after her careful and detailed reconsideration of the role of language during the interview (see pp.75-78).

The following example of an oral exam is especially interesting with respect to separating two assessment processes at the same time in the exam, and highlights the difficulties that arose when the T2 followed her determination to exclude language from assessment. Her effort to assess only the content in the exam led to the paradoxical situation in which she seems to have concentrated on the language

(which she wanted to ignore), and fails to listen to the student's knowledge of the content (which she is expected to assess). This exam situation was very similar to the one at the beginning of my observation (see chapter 6.6), which raised her awareness of the role of language in her assessment in the first place and triggered off her reconsideration. Two students with markedly different linguistic abilities following each other might have alerted her not to repeat the "mistake" of comparing the two students with one another and assigning the second one a lower grade because of his poorer language performance. The situation was exacerbated by the fact that the student who was examined followed the one who had delivered the highly praised performance (see ex. 15).

Example 16:

T: Cuba before the missile crisis and the raise of Fidel Castro
S: Cuba was ruled by Batista since 1940, I think … … …
T: Ja
S: and he allowed the American businessmen to make huge profits in Cuba in gambling casinos and during Mafia (xxx)
T: and who came to these gambling casinos, because that's also an interesting question, who were the customers?
S: ah, … the Cubans
T: no, American business men, this is the ironic thing about it, the Americans came, this was before Castro, and this was, what for instance made Fidel Castro so mad, okay, go on
S: and, and then Fidel Castro rised and he tried to overthrow the government
T: mum
S: but first, mum, versus
T: try, attempt,
S: tried and failed, and was sent into exile
T: mum
S: and then he started guerrilla war and, ah, during this he was able to overthrow the government and, ah, took over ruling. He threw out American business men of the country and, ah, ah, and took over American property
T: ja
S: in amounts of millions (xxx)
T: ja
S: this upset the Americans ...
T: mhm
S: they broke up every contact, as well as econ, econom, economically, and Cuba searched for contacts with the Soviet Union
T: why
S: mhm, for, mhm, ... finding trading partners, for absetzen, dass sie absetzen können ihr Zuckerrohr
T: to find markets for their sugar cane, mhm
S: and the Soviets to establish missile stations on Cuba because the Soviets thought that Cuba was a very good starting point for tactics

T: and also based on the same ideology because both were deeply communist countries and so you take our raw materials and okay come with your missiles and put them there. All this sounds so nice and so friendly. Who didn't like these ideas?
S: the Americans didn't accept this because Cuba was only hundred miles away from America and the rockets, ah, the missiles were able to reach American ground, which wasn't able from the Soviet Union and so they were seeking for destroying the missile stations to protect themselves. The CIA planned a secret attacking the ah the Bay of Pigs, which failed
T: okay, what was the special trick in their plan?
S: they used Cuban exilians
T: exile agents, actually, trained people. Why did they do so, what was the idea behind?
S: they thought these people felt hatred against the actual Cuban ruling, but they were wrong and their attack failed
T:mhm
S: so Kennedy and Khrushev
T: ok, this is part of the next question. Your job was to do the first part, how it came to the invasion in the Bay of Pigs. Why did America actually become so nervous because what we know and always have said is that Khrushev and the Russians concerning their arms were not really so well established, so they could also have said, well, there're a few old fashioned missiles there. What made them so nervous?
S: they thought that the Russians were also able to launch atomic, ah nuclear weapons
T: and what would you say about the geographical area we're talking about, the position, the geographical position?
S: Cuba is very close to America
T: mhm
S: and so überbrücken the distance between
T: bridge the distance
S: bridge the distance between America and Russia without to be afraid that Russian ground will be destroyed
T: ok also very good

Immediately after the exam she commented on the student's performance as follows:

Ja, das war jetzt genau so eine Situation wie damals, da hätte ich dem B. keine so gute Note gegeben, weil C. kann sich halt besser ausdrücken, ja, ihr Englisch ist halt besser, aber die Sprache zählt ja nicht. Das merk ich erst jetzt, also wars gut, dass die Frau Hönig da war. Vor dem Projekt hätte der B. sicher eine schlechtere Note bekommen, aber inhaltlich hat er ja auch alles gesagt was wichtig war.

*//Yes, this was the same situation as then, then I wouldn't have given B. such a good grade, because C. can express herself better, yes, her English is better, but the language doesn't count. I'm only aware of it now, so it was good that Mrs. Hönig was here. Before the project B. would have got a lower grade, but content-wise he also said everything important//*

Her written comment on this exam, which followed some time later, however, spoke a different language:

> Note: Befriedigend. Einige schwere inhaltliche Fehler, müssen von der Lehrerin korrigiert werden, keine eigenständige Korrektur des Schülers. Der Prüfungsdialog ist nur durch Fragen der Lehrerin in Gang zu halten, Frage/Antwort Spiel.
>
> *//Grade: satisfactory. Some severe content errors, must be corrected by the teacher, no self-correction. The dialogue in the exam is only kept going through teacher's questions, question/answer game//*

The transcript of the oral exam enabled the teacher to re-check the student's answers. It is similar to a written exam, in which the teacher can make a selection of what to assess (see chapter 6.2). Nevertheless, the grade she assigned in the written comment differs markedly from her immediate assessment and so does her judgement on content errors. The transcript exhibits only one piece of overt negative feedback – no, no! (see example 8), while the teacher speaks of "several severe mistakes" in the written comment.

What underpins the assumption that this student received the "Sehr gut" only because of the teacher's fear of misjudging him due to his weaker language performance is the fact that she prompted him with long contributions explaining particular situations in the Cuba Crisis. Furthermore, he was the only student who received this grade without having been asked an additional question to do with logic or explaining. That is what the teacher herself did in this case. Usually only those students were awarded a "Sehr gut" who were asked questions requiring higher thinking skills rather than the mere reproduction of facts.

What also differs is the wording of the teacher's feedback in this exam. While all the other students who got a "Sehr gut" for their exams received feedback such as "yes, very nice, J., das macht das Sehr gut aus so zu denken", or "excellent, really, sehr gut", the teacher commented this exam with a mere "okay, also very good", immediately justifying why she awarded the student this grade.

Teachers know the students they assess, they know how much they can expect from them in the exam. It is well-documented that teachers' behaviour in exams mirrors the expectations they hold towards their students.

> […] "Highs" get more and more difficult learning tasks, more feedback and also feedback with a richer content. "Lows" get fewer questions in class, less difficult questions, less opportunity for response and less opportunity to pose questions than the "highs" (Teunissen 1992:98 quoted in Mohan 2001:178).

The question, for example, for the difference between atomic and hydrogen bombs (see ex. 11) might be irrelevant or inappropriate in a history class and might not have been discussed in the lessons. I assume the teacher knew that the student she asked had a sound knowledge of the matter and she offered him the floor to demonstrate it. Not only was he given the chance to get a good grade, he also gained some self-confidence from mastering this task in a foreign language, which his posture showed when he returned to his desk.

**Assessment of written exams**

The teacher contends that correcting written exams allows her to focus on the selection of the criteria she wants her students to fulfil; she can concentrate on the content and pay less attention to the language. A written exam is a document lying in front of her, which can be referred to after the exam; it can be re-read and re-checked: has the student answered all the content questions satisfactorily? She admits, though, that she perceives a difference between correcting linguistic errors and judging the quality of expression: it is easy for her to pass over spelling or grammatical errors but not so easy to do those students justice who show less mastery of language than others.

> Bei schriftlich da ist es wirklich einfach, naja, man kommt schon in Versuchung wie z.B. 3.Person „s", oder die Zeit oder spelling auszubessern, aber das sieht man ja gleich, ich habe das Wort schon wieder ausgebessert, weil er assassination immer mit einem s schreibt oder so– geht mich nichts an. Das schaut man sich ja auch länger an, da korrigiert man das einmal durch, dann überlegt man sich die Punkteanzahl und da kann man sich wirklich gut auf den Inhalt konzentrieren. [...] Da kann man auch wirklich für den Inhalt eben die Punktezahl vergeben, nicht nach irgendwelchen

sprachlichen Sachen. Weil die Stimme weg ist und dadurch die Sprache so weg ist.

Wenn einer sich schriftlich super ausdrücken kann, da denk ich schon ists schwerer, aber da kann man sich immer wieder daran erinnern. Wenn dann da ein paar deutsche Wörter drinnen sind beim andern, dann ist es nicht so eindrucksvoll, aber man kann sich einbremsen, was tue ich da, wieso kriegt der jetzt 3 Punkte, der hat ja alles erwähnt, nur halt in 3 Wörtern und nicht in 2 Zeilen. Da hat man mehr Zeit zum Korrigieren. Man kann sich beim Korrigieren von schriftlichen Sachen besser wieder in Erinnerung bringen, Sprache nicht.

*// with written exams it's really easy, well. you might be tempted to correct the 3rd person s or the tense or the spelling, but you see it immediately, I've again corrected because he always spells assassination with one s - it's none of my business. You look at it longer, you correct it once and then consider the number of points and then you can concentrate on the content really well. Then you can really give the points for the content, not for some language things. Because the voice is gone and with it the language.*

*If someone can express himself excellently then I think it's more difficult, but then you can always remind yourself. If another one uses some German words, then it's not so impressive but you can stop and ask yourself what am I doing here. why does this one get 3 points, he's mentioned everything, only with 3 words and not in 2 lines. You have more time for correcting. with correcting written things you can remind yourself again, language not//*

"The voice is gone and with it the language", the teacher argues. The speech act is gone, indeed. Yet the language remains since the knowledge the students have about the content is demonstrated through the language. The teacher's statement about the exclusion of language from the assessment of written exams poses some challenging questions. However, they do not differ considerably from the questions asked about the role of language in oral exams. To begin with, it should be asked if an understanding of historical events and the ability to express it can be separated at all. Furthermore, do fluency in speech and good writing skills have elements in common? Do good writing skills not have the same effect on students' ability to communicate factual information as does fluency in oral speech? Do students who have the ability to express themselves skilfully in writing not have an advantage over students with difficulties in writing?

What follows from the teacher's statement is the implication that what is not assessed (or is thought to be not assessed?) does not exist or, at least, is not important.

However, if language is unimportant, that is no foreign language production is required, the question should be considered why then the teacher asks students to write short essays about complex historical interrelations (see chapter 6.2). If we can believe what the teacher says, the enumeration of historical events and their important ramifications would suffice. As a consequence, the language used in the written exam would be reduced to listing facts. Nevertheless, in written exams, in which teachers' prompting is absent, I would argue that coherent and logical speech together with the flow of argumentation is indispensable to explain complex historical matters and arguably even more important than in oral exams. .

Furthermore, the teacher's considerations on the role of the language in written exams is a direct contradiction to her comment on the oral exam in example 13 (see above), in which she criticises the student's lack of fluency and small share in the spoken interaction. Yet, in the absence of evidence in the form of corrected written exams, this matter cannot be verified or further discussed.

### 6.7 Language across the curriculum - the paradox continued

Mohan argues that content is inseparable from linguistic expression (Mohan 1986) And yet language is neglected as a medium of learning in subject matter learning and vice versa content in language learning. (see chapter 4.1). This paradox can, as this study shows, be extended to assessment.

Interestingly, this problem is not present in students' minds. For them, the matter presents itself entirely differently: they are well aware that content is embodied in language. In a short discussion with some students from this study about German students' preference for CLIL over language classes (cf. Ernst 1995), there was some kind of agreement among the Austrian students that they do not perceive any difference:

> S: Ist ja egal über was wir reden. In den Englischstunden müssen wir ja auch über was reden, da müssen wir ja auch was wissen. Wir reden ja immer über was.

*//It doesn't matter what we're talking about. In English lessons we also have to talk about something, we also have to know something. We always talk about something//*

One student added in low voice:

najo, wenn ana dauernd olle Fehler korrigiert, dann redn ma net so gern

*//well, if someone is always correcting all the mistakes, then we don't like talking//*

Greater focus on form is undoubtedly attributed to English lessons than to CLIL. Yet in both cases the aim is to enhance English. Perhaps it depends on the individual teacher's skills in either subject to know when and what to correct. What undoubtedly discourages students from speaking freely is the constant interruption of their flow of speech if they make mistakes, no matter whether it is in an English or CLIL class.

Among the teachers in the CLIL classroom, however, the tension between language and content is central. The boxes below show the gap which exists between "theory" and "practice" in three different areas: curriculum planning, exam regulations and teachers' classroom practice:

| theory | practice |
|---|---|
| Assessment begins with the planning of the curriculum | No curriculum for CLIL exists→ no linguistic objectives exist according to which students can be assessed |
| Students are offered the choice of the language in exams | Students do not choose the mother tongue option - "sie sprechen, wie sie es serviert bekommen" |
| Teachers claim not to assess the language | **Language is co-assessed** |

# 7 Summary and discussion of findings

The analysis of the data shows that assessment in CLIL is riddled with discrepancies and contradictions. Among teachers, to begin with, there was extreme reluctance to have their assessment practices observed. To some extent, this unwillingness was not surprising. There is repeated reference in the literature to the fact that assessment is a very private affair and teachers do not want anyone to look over their shoulders when assessing their students' exams.

On the other hand, teachers were willing to give interviews, in which they quite openly answered all the questions asked. Their explicit statements, however, could only be partly verified by actual classroom observation. Such contradictions and also straightforward direct answers to my research questions will be dealt with in more detail in the following.

**Components of assessment**

By and large, the assessment activities observed in this study were carried out in the traditional way by setting oral and written exams. In general, the teachers preferred oral exams but due to lack of time they were sometimes forced to turn to written exams. One teacher (T2), advocated both formats; she thought that oral as well as written exams prepare her students adequately for their future academic education. The teacher, whose lessons I was allowed to observe (T2), had the criteria that had to be fulfilled by the students in oral exams in her mind, whereas she wrote them down prior to written exams. There she formulated the questions in advance and allotted them a certain number of points which students should obtain in order to score full marks. If students did not answer the questions satisfactorily, she then deducted points.

Alternative assessment in form of presentations was carried by two teachers. One teacher (T1) used students' presentation as a form of summative assessment that decides the grade. She set no selective criteria for her students' presentation; what counted as successful exam was whether the students had done the presentation or not. For another (T3), the presentation of a group's summary was part of summative assessment and contributed to the grade the students would receive. However, only

T1 required her students to make their presentations in English, whereas T3's students gave them in German.

**Choice of code**

Students are offered the choice of taking the exam in the mother tongue or the target language (see chapter 3.2). However, students hardly ever availed themselves of the mother-tongue option; they nearly always spoke the language in which the topics had been taught, or to express it in one of the teacher's voices: "sie sprechen, wie sie es serviert bekommen".

There are hardly any exceptions to this behaviour. During my observation only one student preferred to take an exam in German although the topics on which he was examined had been discussed in English. The reason he gave was that he would do history in his school leaving exam in German. In one interview, the teacher spoke of one student who opted for German instead of English for safety reasons. The exam he had to take decided whether he would fail the whole year.

It should, however, not be overlooked that the linguistic competence students have does not always suffice to communicate their content knowledge. For instance, a student of the 5$^{th}$ form resorted to her mother tongue because she wanted to tell the teacher and her peers far more about the topic she presented than she could express in English.

**Code-switching**

If students had vocabulary difficulties in exams, that is to say they did not know the appropriate word in English or could not retrieve it from their memory, they made use of the mother tongue. The word in German, usually in a questioning tone, made the teacher step in with the relevant word in English (see example 3 and 4). Some students repeated the word, others did not, and continued their speech. Even if code-switching was not only intra-sentential but encompassed a longer explanation given in German (see example 5), it was tacitly tolerated. Code-switching was accepted by the teachers as communication strategy in the CLIL classroom and also made no difference to the results of the exams. All the above mentioned examples stem from

exams graded with “Sehr gut” and thus endorse the teachers’ statements that code-switching has no negative affect on the grades.

**Error correction**

There was no consensus among the teachers in this study as to whether to correct linguistic errors or ignore them. Some followed their (assumed ?) duty and corrected errors, others did not want to interrupt the flow of their students’ speech and passed over errors as long as understanding was ensured. It should be noted, though, that the students in class, not the teacher, may well intervene in particular cases when their peers made pronunciation errors, and required their correction (see 6.2).

Yet there was absolute agreement that teachers would not assess linguistic errors, a fact which they inform their students about at the beginning of CLIL instruction. Here they made no difference between pronunciation errors, lexical errors or grammatical errors; none of them affected the grading of the exam. Examples 6, 7 and 8 serve as evidence for the teachers’ assertion that they do not take linguistic errors into account in assessment as they all are taken from exams that were awarded “Sehr gut”.

Since CLIL lessons are content lessons, it goes without saying that content errors are corrected as well as assessed. While I was observing exams, though, students hardly ever made factual errors. If they did not know the answer, they preferred to remain silent.

**The role of language in assessment**

The present investigation of code-switching and error correction in CLIL exams has produced expected as well as provable outcomes: language errors did not matter in assessment, nor did code-switching. The findings concerning the role of the language in assessment, however, were not as straightforward; rather they were conflicting. There was considerable disparity between what the teachers said in the interviews and what actually happened in the classroom. In the interviews they convincingly expressed the view that assessment posed no problems or caused no difficulties for

them, which strongly contradicts what is mentioned in the research literature. Even the question of how to deal with language in assessment, which has been the most contentious issue in the discussion about assessment in CLIL, was answered without hesitation in a straightforward way: the teachers asserted unanimously that they took no notice of the language the students used in their performances, they did not measure and assess the students' language skills but took solely their content knowledge as the basis for their assessment.

Actual observation in class, however, revealed the opposite. It was not only the content knowledge the students showed in their performances that counted, but rather students with greater linguistic skills outperformed those who expressed themselves in poorer language. In other words, students who might have had solid content knowledge but could not put it across eloquently were the "losers" in oral exams. Even though teachers asserted in the interviews that it had been given no weighting, language became the decisive factor in fixing the grades the students were given. This again might not be surprising. Which teacher does not know students who have successfully talked themselves out of predicaments? How many students admit that it was only their eloquence that helped them to survive in exams? It is widely agreed that students with rhetorical skills are more likely to fill the gaps in their knowledge and satisfy teachers than those who are less articulate. What was astounding, though, was the teachers' absolute belief and conviction that they took no account of the language in assessment. This, however, underpins the assumption that there is no general awareness of the linguistic dimension in content subjects both in L1 and L2 teaching (conf. chapter 4.1). The teachers' perception of language is limited to the "superficial" factors, namely the errors the students make in the target language and code-switching. There is, however, no awareness of how fluency and the speed and share of speech affect the way their performance is perceived.

## 7.1 Teaching is not only assessment – the success factors in the context studied

Despite the considerable disparities in assessment that have become obvious in my investigation, my observation also showed how CLIL worked, in particular, how well it worked: 9 out of 19 students of T2's form went in for the school-leaving exam in this particular CLIL subject, and all but one did the exam in English. And they all

passed, as I have learned meanwhile. Moreover, the chairperson of the examining committee of the school-leaving exam, who has some experience as a CLIL teacher, congratulated the teacher on her students' remarkable language performance. In one particular case, for example, it was only the student's language performance that saved him from failing the exam. His content knowledge was not too sound, as the teacher told me, and would possibly not have sufficed for a pass grade. Yet he spoke and spoke, so that his language abilities and skills eventually made up for his lack of content knowledge. The chairperson justified the decision by emphasising that language competence will be of greater benefit in the student's further life than content knowledge.

It could be argued that students at this stage of their education are confident and natural speakers of English, having learned English for at least 12 years at school. In the ideal case this argument is certainly true. Yet ideal cases are rare. Therefore one needs to consider how the students gain their confidence in speaking and their self-evident use of the language.

As mentioned above (see chapter 6.1), I also had the opportunity to observe an English class, in which T2 wanted to show me how she introduced CLIL to the students in lower secondary education in order to make them familiar with learning content in a foreign language. Here I noticed that the lesson began with a lively discussion among the girls about the previous night's episode of a popular TV series, while the boys spoke about recent sports events. It was interesting to watch that it was not the teacher who initiated the discussion but the students began talking by asking the teacher, "Frau Professor, did you watch…?, Frau Professor, do you know…?. The teacher joined in, expressing interest in the complicated relations in the soap as well as the results of sport competitions.

After class T2 told me that she takes special care to create and maintain an anxiety-free climate in the classroom, so she usually reserves the first ten minutes of each English lesson for free talk about topics of interest to her students. Her focus here is on the functional communication among her students and her and the question of making errors or not is of no importance. Her students appreciate this procedure and have become familiar with it. After the discussions she continues with the regular curriculum.

The teacher herself explains the reason why her students become confident, natural speakers of English with the following words:

> Ich denke, dass ein wesentlicher Punkt darin liegt, dass ich von Anfang an die Angst nehme, in der Fremdsprache Fehler zu machen. Mir ist es immer wichtig, zu betonen, dass auch mein Englisch weit davon entfernt ist, perfekt zu sein. Die SchülerInnen wissen, dass ich noch immer ein Vokabelheft führe, dass ich immer wieder versuche, mich zu verbessern. Ich erzähle, wie es in Firmen läuft, dass es darum geht, in der Fremdsprache zu argumentieren, ein Konzept darzustellen, ein Produkt darzustellen, etwas zu verkaufen, etwas zusammenzufassen etc. Und versuche immer zu betonen, das Resultat ist wichtig, die Tatsache, dass der andere mich versteht und ich den anderen. Nach wie vor liebe ich diese Sprache und ich denke, das merken die SchülerInnen. Irgendetwas übernehmen sie, sonst würde es mir nicht passieren, dass immer wieder SchülerInnen auch von der Literatur begeistert sind, die ich vorschlage, ich denke nicht, dass dies passiert, weil sie mir einen Gefallen tun wollen.
>
> Das Gleiche passiert in Geschichte: sie wissen, wie wichtig mir dieser Gegenstand ist und ich lade sie dazu ein, gleichzeitig eine Fremdsprache zu üben. Das übernehmen sie. So hat sich zum Beispiel ein sehr begabter Schüler meiner jetzigen 8ten Klasse mit ein bisschen Überredungskunst meinerseits an ausländischen Unis beworben, mit einem Referenzschreiben meinerseits (an dem ich stilistisch lange gefeilt habe) und wurde kurz vor Weihnachten von Oxford (!) zu hearings eingeladen. Habe dies (natürlich unbezahlt) mit ihm geübt, das sind dann meine " Highlights".
>
> *//I think a crucial point is that I take away their fear of making mistakes in the foreign language right from the beginning. It is always important for me to emphasise that my English is far from being perfect, too. The students know that I still keep a vocabulary list, that I always try to improve my English. I tell them how it works in companies, that the point is to bring forward your arguments in the foreign language, to present a concept, to present a product, to sell something, to summarise something, etc. I always try to emphasise that the result is important, the fact that the other one can understand me and I*

*can understand the other one. I still love this language and I think that's something my students realise. They respond to that positively, otherwise it wouldn't happen to me that my students also like the reading I suggest, and I don't think it only happens because they want to please me.*

*The same happens in history: they know how important this subject is to me and I invite them to practice a foreign language at the same time. That's what they take on board. So it happened that a very gifted student from my current 8th form, after some persuasion on my part, applied to foreign universities, with a letter of reference I wrote for him (which I worked on stylistically for quite some time) and he was invited to interviews at Oxford (!) shortly before Christmas. I've practiced this with him (of course without pay), these are then my "highlights"//*

The teacher attributes her students' language competence in both the English classes and CLIL to the anxiety-free atmosphere that prevails in the classroom and her focus on the functionality of the language. She allows her students to make mistakes without being sanctioned and puts across to them the idea that it is the message that decides if communication is successful or not. Furthermore, the teacher manages to convey to her students the passion she feels for her subjects. What I witnessed throughout my observation was an atmosphere of mutual esteem and a teacher who enjoyed teaching. Put in a nutshell, CLIL worked so well due to operational proficiency on the part of the students and great commitment on the part of the teacher (but is not the latter the reason why teaching works well in general?).

I would like to argue that an additional factor might have contributed to the fact that T2's students succeeded in the school-leaving exam. If we look at the students' individual preparation for the exam, we can see that there are close similarities to portfolio work. The steps the students take on their way to the exam are the collection of information on the topic they have chosen and the critical selection of material. Although the teacher's individual feedback and support was beyond the score of my observation, I noticed that the teacher advised her students about, for example, how to narrow down the scope of their topics or where they might need or could find further information. What I also noticed was that some students exchanged and discussed their ideas and thoughts about the topics they would

explore. They communicated with their peers in their mother tongue, though, but nevertheless they had to work with the target language while preparing for the exam. In the exam they then had to present their knowledge.

Portfolio work, however, is not very popular with the teachers I interviewed during my investigation. Their objection raised to this suggestion is predominantly the amount of time that has to be devoted to this kind of work. In Austrian schools, there is traditionally a rigid distinction between main subjects and minor subjects. The former, such as mathematics, German and foreign languages are generally characterised by two tests a term, which take up the lion's share of students' time reserved for studying. The latter, which encompass content subjects like history, have less weight and would therefore not lend themselves for time-consuming tasks such as portfolios (oral communication).

## 7.2 Reflecting the research process

The journey I embarked upon to find out how assessment in CLIL was carried out in practice might well serve as an explanation or an answer to the question why assessment in CLIL is frequently referred to as being problematic. It was long and rough and closely mirrored the difficulties mentioned in the literature: teachers strongly object to having anyone look at their assessment procedures; they prefer to keep their practices as well as their experiences to themselves, and do not communicate them to their colleagues. The teachers' reluctance to make assessment a "public affair" might not be attributed solely to CLIL; assessment is a delicate matter in general, which teachers prefer to keep to themselves.

The privacy that teachers did not want to be invaded and the lack and breakdown of communication, I had to face at the outset, foreshadowed the reasons why assessment in CLIL is still an unsolved problem. "Access to the field is often not only a technical problem. The difficulties and obstacles that researchers encounter here are frequently already suggestive of a considerable part of the discoveries that may be made about the particular field and the actors concerned, and also touch on the specific role of the researcher in the field" (Flick, Kardoff, Steinke 2004:195). Privacy and lack of communication seem to be constant, and obviously inseparable,

companions to assessment. The researcher, consequently, appears as the interloper, who enters this privacy with no right to be there and might unmask hidden truths.

Specifically, what made my research so demanding and what also explains why both interviewing the teachers and video-taping actual classroom assessment was necessary, was my intention to trace how teachers respond to the dual focus on content and language when assessing their students' performance. I had to overcome several obstacles until I eventually obtained tangible results. At the beginning my endeavour was endangered as it was extremely difficult to find teachers willing to co-operate (see chapter 5.1). Some teachers I asked denied co-operation immediately when confronted with the investigation of assessment. Others were ready to give interviews but adamantly refused to be video-taped. Had I, however, only followed what the teachers told me in the interviews, the results of my research would have been entirely different. What they said they did was one truth; another one, namely quite the opposite, was what actually happened in practice.

After a series of fruitless attempts I found one teacher (T2), who not only offered co-operation but also showed intense interest in my project and, additionally, provided me with active support to get it going. She persuaded a colleague (T1) to take part in my study and won over her students when they then protested against video-taping.

T1 allowed me to video-tape three lessons of her 5$^{th}$ form at the end of the term, which she had reserved for assessment. Yet her assessment practice did not prove useful for investigation in any respect as she accepted her students' scripted presentations, which were taken as the basis for summative assessment, without correction or feedback.

As mentioned before, T2 met my research project with full approval, giving two reasons for her interest: firstly, since no communication among CLIL teachers exists, she hoped to gain information on how her colleagues do assessment in CLIL, and secondly, she generally thinks that external evaluation fosters her personal development. Her readiness, however, was seriously challenged when at the very beginning of my observation she realised that her assessment practice completely contradicted to what she had told me shortly before: not only did she assess the language, which she, and with her all the teachers I interviewed, had strongly denied

previously, but the students' linguistic performance also decided the grade they received.

Unfortunately, this unexpected outcome of my very first encounter with assessment in CLIL led to a breakdown of communication between T2 and me. To my understanding, it took her some time to "digest" the fact that she did something she was firmly convinced that she would never do. What might have made matters worse was the fact that she had also assured her students that the language would never be taken into account in her assessment. In this context, it needs to be mentioned that the teacher enjoys a high reputation as an excellent teacher among parents and students; in particular her students acquire an outstanding competence in speaking English.

However, it should not be overlooked that this experience led to positive results on the teacher's side. It made her reconsider her assessment practice and, particularly, it raised her awareness of the role the language plays in performance, as she later on told me in an extended interview. Her effort, however, to take special care in future to concentrate only on content and leaving language out of consideration in assessment ended unsuccessfully. Further observation in class and the teacher's comments on the transcribed exams showed that, in particular in oral exams, it was impossible to separate content from language.

### 7.3 Possible solutions

The investigation of oral exams revealed that no validity is given: to a great extent the students are not judged on the content of the subject but on the language they produce in their performance. The teacher herself came up with a suggestion in order to minimize the influences in an oral exam and thus increase its validity. She suggested that the scope of the exam, listing all the questions she wants to have answered, should be defined and written down prior to the exam (conf.2.5). During the exam she can then control whether the questions have been fully answered or not. Thus, students with less language proficiency would get the same chances as eloquent students.

> Eigentlich müsste man sich vorher aufschreiben was man will, nur stichwortartig. Und dann Hakerl, Hakerl, Hakerl, das hat er genannt und das auch, ist ein Sehr gut. Weil sonst lässt man sich von der Sprache verleiten.
>
> *//You ought to write down in advance what it is you want to have mentioned, only key points. And then tick, tick, tick, he's mentioned this and that, is a "Sehr gut". Otherwise you're misled by the language//*

Her suggestion perfectly ties in with what Neuweg (2006:47/48) recommends for the avoidance of the negative influences in oral exams which skew the grades (see chapter 2.5). Such problems, though, cannot be ascribed solely to CLIL. They occur in all exams that have language as their medium and thus cast serious doubt on their validity.

This, however, is only one side of the coin, referring to oral exams in general. The other side of the coin is that CLIL is an educational approach whose aim is twofold: to promote subject knowledge and simultaneously enhance language competence. Whereas the learning objectives in the content area of CLIL are defined according to the curriculum of the respective subject in the mother tongue, the role of language is overlooked on the level of the curriculum; no attention is paid to learning objectives in the language field. Researchers and educationalist agree that language in CLIL must be seen from its functional perspective so that relevant language objectives can be defined in the curriculum. Yet, even if such a school curriculum for CLIL existed, there would still be a number of practical problems with regard to the language in the process of assessment. For example, agreement needs to be reached on the weighting of language in assessment. Against which criteria or reference norms should students be assessed (see chapter 2.2)? If internal norms are used, the grade has a certain value only within the classroom. Should external criteria, such as established in the CEF, be used as a basis for assessment? Which criteria could provide the basis for CLIL diplomas, whose introduction is suggested in the CLIL compendium (see chapter 3.2)? The number of questions and their complexity indicate that there might be a long way to go for CLIL to "mature" as well as to find satisfactory answers in the area of assessment.

## 8 Conclusion

My investigation shows that it is very difficult, or even impossible, to separate content and language in assessment. This fact needs to be recognised and programmatic statements in this respect need to be removed from CLIL guidelines. What follows is that established assessment practices are incapable of measuring students' achievements satisfactorily. Surely, this indicates that it would be necessary to look around for assessment instruments that incorporate both content- and language-focused criteria and help students to get due credit for the knowledge and skills they demonstrate. The development of these instruments is both a challenge and an invitation to practitioners, educationalists and researchers to raise the language awareness among teachers and to define its role in learning, and consequently in assessment.

The opera *Capriccio* has the competition between music and words at its core. What better comment on the inseparability of words and music – or content and language – could there be:

Vergebliche Müh'n
die beiden zu trennen.
In eins verschmolzen
sind Worte und Töne.

*Love's labour lost*
*when trying to separate.*
*Intimately woven into a unity*
*are words and sounds.*
(my translation).

## 9 Bibliography


Abuja, G.; Heindler, D. und Projektgruppe „English als Arbeitssprache“. (Hrsg.). 1995a. Lehrplananalyse. EAA Servicehefte 1 Graz: ZSE, Bereich III.

Abuja, G.; Heindler, D. und Projektgruppe „Englisch als Arbeitssprache“. (Hrsg.). 1997. Paxisberichte. EAA Servicehefte 4 Graz: ZSE, Bereich III.

Abuja, Gunther (Hrsg.). 1996. Fremdsprachen als Arbeitssprachen: die Rolle des “native speakers”; Projektgruppe „Englisch als Arbeitssprache“ Graz; Wien: Zentrum für Schulentwicklung.

Abuja, Gunther; Simpson, Stuart. 1997. Bericht über den Europarat Workshop 12B. Bilingual Education in secondary schools: learning and teaching non language subjects through a foreign language. Graz: Zentrum für Schulentwicklung. Bereich III: Fremdsprachen.

Abuja, Gunther (Hrsg.). 1998. Englisch als Arbeitssprache: Modelle, Erfahrungen und Lehrerbildung. Graz: Zentrum für Schulentwicklung. Bereich III, (Berichte/ Bundesministerium für Unterricht und kulturelle Angelegenheiten, Zentrum für Schulentwicklung, Bereich III: Reihe 3).

Abuja, Gunther and Dagmar Heindler (Hrsg.). 1993. Englisch als Arbeitssprache: Fachbezogenes Lernen von Fremdsprachen. Graz: Zentrum für Schulversuche und Schulentwicklung.

Abuja, Gunther. 1999. Die Verwendung einer Fremdsprache als Arbeitssprache: Charakteristika bilingualen Lernens in Österreich. Zs für interkulturellen Fremdsprachenunterricht. http:www.spz-tu-darmstadt.de/project_ejournal/ig-04-2/beitrag/abuja2.htm (7 May 2007).

Aust, Stefan, Claus Richter, Matthias Ziemann. 2008. *Wettlauf um die Welt: Die Globalisierung und wir*. München: Piper.

Bach, Gerhard and Susanne Niemeier (Hrsg.).*Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* Frankfurt: Peter Lang.

Bachmann, Lyle.1990. *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Bausch, Karl-Richard; Christ, Herbert; Krumm, Hans-Jürgen (Hrsg.). 2003. *Handbuch Fremdsprachenunterricht.* Tübingen und Basel: Francke.

Birenbaum, M. 1996. Assessment 2000: towards a pluralistic approach. In M. Birenbaum & F.J.R.C. Dochy (eds.). *Alternatives in Assessment of achievements, learning processes an prior knowledge* (pp. 3-29). Boston, Dordrecht & London: Kluwer Academic Publishers.

Brinton, Donna, Marguerite Anne Snow and Marjorie B. Wesche. 1989. *Content-based second language instruction.* New York: Newbury House.

Butzkamm, Wolfgang. 1988. Code-switching in a Bilingual History Lesson: The Mother Tongue as a Conversational Lubricant. http://www.multilingual-matters.net/beb/001/0081/beb0010081.pdf (7 June 2008).

*CLIL Compendium* www.clilcompendium.com

*Content-English* www.content-english.org

Dalton-Puffer, Christiane. 2002. Content and language integrated classrooms in Austria: applied linguistics takes a look. VIEWZ 11/1, 4-26. http://www.univie.ac.at/Anglistik/ang_new/online_papers/views/Views11b.pdf (2 Feb. 2008).

Dalton-Puffer, Christiane. 2003. Telling each other to do things in class: directives in content and language integrated classrooms. VIEWZ 12/1, 3-23. http://www.univie.ac.at/Anglistik/ang_new/online_papers/views/03_1/views031.pdf (2 Feb. 2008).

Dalton-Puffer, Christiane. 2004. Academic language functions in content and language integrated classrooms: defining and hypothesizing. VIEWZ 13, 23-48. http://www.univie.ac.at/Anglistik/ang_new/online_papers/views/04_1/0401ALL.pdf (2 Feb. 2008).

Dalton-Puffer, Christiane. 2005. *Immersion classrooms as language learning environments : applied linguistic perspectives*. Habilitationsschrift. Universität Wien.

Dalton-Puffer, Christiane; Ute Smit (eds). 2007. *Empirical perspectives on CLIL classroom discourse.* Frankfurt am Main; Wien (u.a.): Lang.

*Dalton-Puffer, Christiane. 2008.* Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe. In Werner Delanoy and Laurenz Volkmann (eds.) *Future Perspectives for English Language Teaching*. Heidelberg: Carl Winter. Manuscript.

Echevarria, J. M., Vogt and D. Short. 2004. *Making Content Comprehensible for English Learners.* New York: Pearson.

Edmondson, Willis. 1986: Some way in which the teacher brings errors into being. In: Kaspar, Gabriele (ed) 1986. *Learning, teaching and communication in the Foreign Language Classroom.* Aarhus: Aaarhus University Press.

Ellis, Rod: 1997. *The study of second language acquisition.* Oxford: Oxford Univ. Press.

Ernst, Manfred. 1995. Fehlerkorrektur und Leistungsbewertung im bilingualen Sachfachunterricht. In Praxis des neusprachlichen Unterrichts, 42 /3, 258-264.

European Commission. 2002. *CLIL/EMILE The European Dimension*. Jyväskylä: University of Jyväskylä.

Fehling, Sylvia; Claudia Finkbeiner. 2002. „Evaluation von Schulleistungen im bilingualen Sachfachunterricht“. Bilingualität und Mehrsprachigkeit. Modelle, Projekte, Ergebnisse. Claudia Finkbeiner (Hrsg.). Perspektiven Englisch Heft 3. Hannover: Schroedel. 22-32.

Finkbeiner, Claudia (Hrsg.) 2002. *Bilingualer Unterricht. Lehren und Lernen in zwei Sprachen*. Hannover: Schroedel.

Fleming, Michael. 2006. Evaluation and Assessment. Languages of Schooling: towards a framework for Europe. Language Policy Division, Strasbourg. http://www.coe.int/T/DG4/Linguistic/Source/Fleming_Evaluation_final_EN.doc (6 June 2008).

Fleming, Michael. 2007. Report of the intergovernmental conference: Languages of schooling within a European framework for languages of education: Learning, teaching, assessment. http://www.coe.int/T/DG4/Linguistic/Source/Prague2007_ConferReport_EN.doc (6 June 2008).

Flick, Uwe, Ernst von Kardorff and Ines Steinke (Hrsg.). 2000. *Qualitative Forschung. Ein Handbuch.* Rowohlts Enzyklopädie. Reinbeck bei Hamburg: rororo.

Flick, Uwe, Ernst von Kardorff and Ines Steinke (eds.) 2004. *A Companion to Qualitative Research. London: Sage Publications.*

Freitag, C.& C. Solzbacher (Hrsg.). *Anpassen, Verändern, Abschaffen? Schulische Leistungsbewertung in der Diskussion*. Bad Heilbrunn: Klinkhardt.

Fruhauf, Gianna, (Hrsg.) 1996. *Fremdsprache als Arbeitssprache in Sachfächern: Praxis und Perspektiven bilingualen Lehrens und Lernens in Europa.* Alkmaar: Stihting Europees Platform voor het Nederlandse Onderwjis.

Gierlinger, Erwin. 2002. Fremdsprache als Arbeitssprache in Österreich: Situation und Perspektive (Comenius 2.1.project) http://www.pze.at/memo (18 April 2007).

Grangl, Sylvia. 1998. In Englisch als Arbeitssprache: Modelle, Erfahrungen und Lehrerbildung. Graz: Zentrum für Schulentwicklung. Bereich III, (Berichte/ Bundesministerium für Unterricht und kulturelle Angelegenheiten, Zentrum für Schulentwicklung, Bereich 3: Reihe 3).

Hallet, Wolfgang. 1998. The Bilingual Triangle. Überlegungen zu einer Didaktik des bilingualen Sachfachunterrichts. In Praxis des neusprachlichen Unterrichts, 25/2, 115-125.

Harley, Birgit, Patrick Allen; Jim Cummins, Merill Swain. 1990. *The development of second language proficiency.* Cambridge: Cambridge University Press.

Häusler-Greenfield, Pat; Glyn Hughes; Ron Ritchie (eds.). 1999. *Learning through the Romeo experience.* The Romeo Project dissemination pack. European Commission.

Helbig, Beate. 2003. Bilinguales Lehren und Lernen. In: *Handbuch Fremdsprachenunterricht.* Tübingen und Basel: Francke.

Heller, Kurt A. 1984. *Leistungsdiagnostik in der Schule*. Bern Stuttgart Toronto: Huber.

Ingenkamp, K. 1989. *Diagnostik in der Schule. Beiträge zu Schlüsselfragen der Schülerbeurteilung.* Weinheim: Beltz.

Ingenkamp, K. (Hrsg). 1977. *Die Fragwürdigkeit der Zensurengebung.* Weinheim und Basel: Beltz.

Ingenkamp, K. 1997. *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.

Jäger, Reinhold S. 2000. *Von der Beobachtung zur Notengebung. Ein Lehrbuch.* Landau: Empirische Pädagogik.

Johnson, R. K, and M. Swain (eds.). 1997. *Immersion Education: International Perspectives.* Cambridge: Cambridge University Press.

Jürgens, Eiko. 2000. *Leistungserziehung und Leistungsbeurteilung*. Neuwied: Luchterhand.

Kiely, Richard. 2009. CLIL – The question of assessment. http://www.developingteachers.net/articles_tchtraining/clil1_richard.htm (6 Jan 2009).

Klingler, Josef. 1982. *Objektivität und Gerechtigkeit bei der Beurteilung von Schülerleistungen.* Klagenfurt: Universitätsverlag Carinthia.

Krashen, Stephen D. 1982. *Principles and practice in second language acquisition.* Oxford: Pergamon.

Krashen, Stephen D. and Tracy D. Terrell. 1984. *The natural approach: language acquisition in the classroom*. Oxford: Pergamon.

Krumm, Hans-Jürgen. 1990. „Ein Glück, dass Schüler Fehler machen“ Anmerkungen zum Umgang mit Fehlern im lernerorientiertem Fremdsprachenunterricht. http://books.google.at/books?id=_tJdjuiS7UkC&pg=PA128&lpg=PA128&dq=krumm+fehler+machen&source=web&ots=2h6CW0QtBk&sig=aZQSZOy-Y-PNxsAEmohONKNAyUs&hl=de&sa=X&oi=book_result&resnum=7&ct=result#PPA128,M1 (3 Sept. 2008).

Lyle, F., Bachmann. 1990. *Fundamental Considerations in Language Testing.* Oxford University Press.

Mager, Robert, F. 1977. *Lernziele und Unterricht.* Weinheim und Basel: Beltz.

Markee, Numa. 2000. *Conversation Analysis.* Mahwah, NJ: Erlbaum.

Marsh, David. 2002. The relevance and potential of content and language integrated learning (CLIL) for achieving MT+2 in Europe http://web.fuberlin.de/elc/bulletin/9/en/marsh.html (2 June 2008).

Marsh, D. and G. Langé (eds.). 1999. *Implementing Content and Language Integrated Learning*. Jyväskylä: TIE-CLIL.

McNamara, T, F. 1996. *Measuring Second Language Performance.* London: Longman.

Mohan, Bernard; Leung, Constant; Davison, Christ. 2001. *English as a Second Language in the Mainstream: Teaching, Learning and Identity.* Harlow: Pearson.

Morgan, Carol. 2006. Appropriate language assessment in content and language integrated learning. Language Learning Journal, No 33. 59-67.

Neumüller, Josef. 2002. Lebenslanges Lernen als Ziel: Maßnahmen zur Förderung der Bildungsmotivation. In Erziehung und Unterricht 9-10, pp. 1268-1270. Wien.

Neuweg, Georg Hans. 2006. *Schulische Leistungsbeurteilung. Rechtliche Grundlagen und pädagogische Hilfestellungen für die Schulpraxis.* Linz: Trauner – Univ.Verlag.

Nezbeda, Margarete. 2005. Überblicksdaten und Wissenswertes zu Fremdsprache als Arbeitssprache. EAA Servicehefte 6. Graz: ZSE, Bereich III. www.sprachen.ac.at/download/serviceheft5.pdf (5 May 2007).

Olechovsky, Richard and Elisabeth Persy. 1986. *Fördernde Leistungsbeurteilung.* Wien-München: Jugend und Volk

Pfeifer, Silvia, Joachim Kribel. 2007. *Lernen mit Portfolios.* Neue Wege des selbstgesteuerten Lernens. Göttingen: VandenhoeckRuprecht.

Poisel, Eva. 2007. Assessment Modes in CLIL to Enhance Language Proficiency and Interpersonal Skills. VIEWZ 16/3, 43-46 http://www.univie.ac.at/Anglistik/ang_new/online_papers/views/0703.pdf (31 Aug. 2008).

Rheinberg, Falko. 2002. Bezugsnormen und schulische Leistungsbeurteilung. In: Weinert, Franz E. (Hrsg.): *Leistungsmessungen in Schulen*, 2. Aufl. Weinheim: Beltz . 59-71.

Rowntree, Derek. 1981. *A dictionary of education.* London: Harper&Row.

Rowntree, Derek. 1994. *Assessing Students. How shall we know them.* London: Kogan Page.

Schaub, Horst; Karl G. Zenke. 2007. *Wörterbuch Pädagogik.* München: Deutscher Taschenbuchverlag.

Scheiflinger, Werner; Gottfried Petri.1999. *Probleme der Lernerfolgsfeststellung.* Graz: Zentrum für Schulentwicklung.

Schleicher, Andreas. Rohstoff Bildung. In Aust, Stefan, Claus Richter, Matthias Ziemann. 2008. *Wettlauf um die Welt: Die Globalisierung und wir*. München: Piper.

Schratz, Michael. 1996. *Gemeinsam Schule lebendig gestalten.* Weinheim und Basel: Beltz.

Schröder, Hartwig. 1997. *Leistung in der Schule.* München: Arndt.

Solzbacher, Claudia and Christine Freitag. 2001. *Anpassen, Verändern, Abschaffen.* Bad Heilbrunn: Julius Klinkhardt

Thürmann, Eike. 2000. Eine eigenständige Methodik für den bilingualen Sachfachunterricht? In Bach, Gerhard and Susanne Niemeier (Hrsg.). *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* Frankfurt: Peter Lang.

Verordnung über die Leistungsbeurteilung in Pflichtschulen sowie mittleren und höheren Schulen (Leistungsbeurteilungsverordnung). http://www.bmukk.gv.at/schulen/recht/gvo/lb_vo.xml#15 (15 Jan. 2008).

Vierlinger, Rupert. 1999. *Leistung spricht für sich selbst. "Direkte Leistungsvorlage" (Portfolios) statt Ziffernzensuren und Notenfetischismus.*. Heinsberg: Dieck.

Vollmer, Helmut J..2000. Bilingualer Sachfachunterricht als Inhalts- und als Sprachlernen. In Bach, Gerhard and Susanne Niemeier (Hrsg.). *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* Frankfurt: Peter Lang.

Vollmer, Helmut J..2001. Leistungserhebung und Leistungsbemessung im bilingualen Sachfachunterricht: Praxis, Probleme, Innovationen. In C. Freitag,.& C. Solzbacher (Hrsg.). *Anpassen, Verändern, Abschaffen? Schulische Leistungsbewertung in der Diskussion*. Bad Heilbrunn: Klinkhardt. 207-235.

Vollmer, Helmut J..2006. Language across the curriculum. www.coe.int/lang (15 Jan.2008).

Wald, Staudinger, Scheucher, Scheipl, Ebenhoch. 2006. *Zeitbilder 7&8. Geschichte und Sozialkunde, Politische Bildung. Vom Ende des Ersten Weltkrieges bis in die Gegenwart.* Wien: öbvhpt.

Weinert, Franz. E. (Hrsg.). 2002. *Leistungsmessungen in Schulen.* 2. Aufl. Weinheim: Beltz.

Wildhage, Manfred; Otten, Edgar. 2003. *Praxis bilingualen Unterrichts.* Berlin: Cornelsen.

Winter, Felix. 2004. *Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen*. Hohengehren: Schneider.

Wolff, Dieter. 1996. Bilingualer Sachfachunterricht: Versuch einer lernpsychologischen und fachdidaktischen Begründung. http://www2.uni-wuppertal.de/FB4/bilingu/bilingualer_sachfachunterricht.htm (6 June 2008).

Wolff , Dieter. 2002. Fremdsprachenlernen als Konstruktion: Einige Anmerkungen zu einem viel diskutierten neuen Ansatz in der Fremdsprachendidaktik. http://www.babylonia-ti.ch/BABY402/PDF/wolffint.pdf (6 June 2008).

Wolff , Dieter. 2005. Approaching CLIL. Report of Central Workshop 6/: CLIL Quality Matrix. http://www.ecml.at/documents/reports/wsrepD3E2005_6.pdf (12 May 2008).

Wolff , Dieter. 2007. Some Educational and Methodological Principles of Content and Language Integrated Learning (CLIL). http://www.goethe.de/ges/spa/dos/ifs/met/en2747826.htm (12 May 2008).

Ziegenspeck, J. 1999. *Zensur und Zeugnis in der Schule.* Handbuch. Bad Heilbrunn: Klinkhardt.

# APPENDIX

## Transcription conventions

Transcription conventions are a simplified version of the conversation analytic conventions put forward by Atkinson and Heritage 1984: ix-xvi, adapted by Markee (2002:166-167).

**Identity of speakers**

| | |
|---|---|
| T | teacher |
| T1 | identified teacher |
| S | student |
| S1 | identified student |

**Characteristics of speech delivery**

| | |
|---|---|
| ……. | short pause, long pause; within or between utterances |
| ? | rising intonation, not necessarily a question |
| ! | strong emphasis with falling intonation |
| no- | a hyphen indicates an abrupt cut-off |

**Commentary in the transcript**

| | |
|---|---|
| ((laughs)) | comment about actions, including non-verbal actions |
| (xxx) | indicates a stretch of talk unintelligible to the analyst |
| //thanks// | indicates the English translation of an utterance made in German |
| **could you please** | bold font shows material which is currently under discussion |

# LANDESSCHULRAT FÜR NIEDERÖSTERREICH

3 09 St, Pölten, Rennbahnstraße 29 Parteienverkehr Dienstag 8 - 1 2 Uhr

Telefon 02742 / 280 - Klappe e-mail: office@ lsr-noe.gv.at Telefax 02742/280 - 1 1 1 1

Frau

Ingrid Honig
Laubegasse 1 5
2700 WrNeustadt

Präs.-420/256-2007

Bei Antwort bitte Zahl angeben Beilage(n)

| Bezug | Bearbeiterin | Klappe | Datum |
|---|---|---|---|
| | Mag. Unterberger | 5370 | 07.05.2007 |

**Betrifft**

Genehmigung der Durchführung einer empirischen Untersuchung

Der Landesschulrat für Niederösterreich genehmigt die Durchführung der vorgelegten empirischen Untersuchung in Form von Videoaufnahmen zum Thema „Leistungsfeststellung im CLIL - Unterricht“ durch Frau Ingrid Hönig.

Die Untersuchung darf in dem vorliegenden Umfang an der im Antrag angeführten allgemein bildenden höhere Schule durchgeführt werden.

Auf die Einhaltung der Datenschutzbestimmungen darf hingewiesen werden, außerdem ist vor Beginn der Videoaufnahmen das Einverständnis der Eltern bzw. Erziehungsberechtigten und die Zustimmung der Direktion einzuholen.

Für den Amtsführenden Präsidenten

Dr. Freudensprung

Wirkl. Hofrat

Für die Richtigkeit der Ausfertigung

## Interview guideline

1. wie kommen Sie auf die Geschichtsnoten (Inhalt – Sprache)?
2. mündliche, schriftliche Prüfungen?
   unterschiedliche Prüfungen?
3. wo sind Schüler besser?
4. wie oft gibt es Prüfungen?
5. welches Gewicht haben schriftliche, mündliche Prüfungen?
6. wechseln die Schüler die Sprache?
7. wie viele Deutsch, wie viele Englisch?
8. was beobachten sie, wo tun sie sich leichter?
9. fallen Ihnen Schüler auf, die hervorstechen: positiv – negativ?
10. glauben Sie, dass Sie sich im Unterricht auf Deutsch anders verhalten bei der Beurteilung?
11. gibt es etwas, was ich Sie hätte noch fragen sollen?

wie lange unterrichten Sie schon

wie lange CLIL

**Lebenslauf**

Name: Ingrid Hönig, geb. Egger

Geburtsdatum: 14. März 1944

Staatsbürgerschaft: Österreich

Familienstand: verheiratet, 3 Töchter

Beruf: Lehrerin für Englisch in der Erwachsenenbildung seit 1989 (EFL, ESP)

Schule und Ausbildung:

1950-1951 Volksschule Wörth

1951-1954 Volksschule Wimpassing

1954-1962 Bundesrealgymnasium Neunkirchen

1962-1963 Aufenthalt in England

1963-1966 Studium an der Universität Wien:

Englisch
Geographie

1963 First Certificate in English (FCE)

1984 Certificate of Proficiency in English (CPE)

1998-1999 Lehrgang "Lust auf Sprache" am Bundesinstitut für Erwachsenenbildung

2003 TELC – Prüferinnenzertifikat

2005 Comenius Stipendium für „Advanced Language, Materials and Methodology" am NILE Institut in Norwich

2001-2009 Studium an der Universität Wien:

Anglistik und Amerikanistik

Sprachen: Deutsch, Englisch, Italienisch, Französisch